

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 19:52:25 ; Search time 4776.29 Seconds

(without alignments)  
11254.112 Million cell updates/sec

Title: US-08-876-132-2

Perfect score: 1847

Sequence: 1 ACATCTCAACACAGTTTAA.....AAACAGCAAAAGACATCT 1847

## Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Genemdb1:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_in:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1066.6	57.7	3661	1	AF128889	AF128889 Pantoea c
2	130	7.0	7597	6	AX345915	AX345915 Sequence
3	123.2	6.7	6286	6	AX251345	AX251345 Sequence
4	114.2	6.2	6109	6	AX251772	AX251772 Sequence
5	114.2	6.2	6109	6	AX345228	AX345228 Sequence
6	113.8	6.2	11861	9	AC069435	AC069435 Homo sapi
7	113.2	6.2	7814	6	AX251284	AX251284 Sequence
8	112.6	6.1	171105	3	AC007532	AC007532 Drosophila
9	112.6	6.1	190801	3	AC009212	AC009212 Drosophila
10	111.6	6.0	12237	6	AX347260	AX347260 Sequence
11	111	6.0	294914	3	AE003603	AE003603 Drosophila
12	110	6.0	67970	3	PFMAL1P3	AL031746 Plasmid
13	109	5.9	14867	3	AE001398	AE001398 Plasmid
14	108.8	5.9	90487	9	AL592166	AL592166 Human DNA
15	106.8	5.8	213056	9	AL103590	AL103590 Homo sapi
16	105.2	5.7	161078	9	AP003500	AP003500 Homo sapi
17	105.2	5.7	161286	9	AC025120	AC025120 Homo sapi
18	103.4	5.6	100925	9	AC012627	AC012627 Homo sapi
19	103.4	5.6	120481	9	AC099738	AC099738 Homo sapi
20	103.4	5.6	177293	2	AC069525	AC069525 Homo sapi
21	103.4	5.6	235532	9	AC008739	AC008739 Homo sapi
22	103.2	5.6	79018	9	AL627107	AL627107 Human DNA
23	102.6	5.6	151032	9	AC105201	AC105201 Homo sapi
24	102.4	5.5	262608	2	AC092362	AC092362 Homo sapi
25	101.8	5.5	131682	9	AL672277	AL672277 Human DNA
26	101.8	5.5	189082	9	AL731537	AL731537 Human DNA
27	101.2	5.5	349960	6	AX344535	AX344535 Sequence
28	100.8	5.5	170141	2	AL611928	AL611928 Homo sapi
29	100.8	5.5	201312	2	AC067900	AC067900 Homo sapi
30	99.8	5.4	50612	9	AL513530	AL513530 Human DNA
31	99.8	5.4	81696	9	AF438327_3	Continuation (4 of
32	99.8	5.4	86827	3	PFMAL3P5	AL034556 Plasmid
33	99.8	5.4	110000	9	AF438327_2	Continuation (3 of
34	99.6	5.4	17443	3	AC024806	AC024806 Caenorhab
35	99.6	5.4	150119	2	AL845366	AL845366 Danio rer
36	99	5.4	155702	9	AC105451	AC105451 Homo sapi
37	98.4	5.3	136098	9	AC006970	AC006970 Homo sapi
38	98.4	5.3	170141	2	AL611928	AL611928 Homo sapi
39	98.2	5.3	148525	2	AC016675	AC016675 Homo sapi
40	97.8	5.3	13321	6	AX251175	AX251175 Sequence
41	97.8	5.3	158950	9	AL391425	AL391425 Human DNA
42	97.6	5.3	86975	2	AC048340	AC048340 Homo sapi
43	97.4	5.3	162445	9	AL158151	AL158151 Human DNA
44	97.2	5.3	124057	9	AP000770	AP000770 Homo sapi
45	97	5.3	5413	6	AX251450	AX251450 Sequence

## ALIGNMENTS

RESULT 1  
AF128889  
LOCUS  
DEFINITION Pantoea citrea strain ATCC31623 plasmid pZG500, complete plasmid  
ACCESSION AF128889  
VERSION AF128889.1 GI:4457229  
KEYWORDS  
SOURCE Pantoea citrea.  
ORGANISM Pantoea citrea.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
- Pantoea.  
REFERENCE 1 (bases 1 to 3661)  
AUTHORS Billic.M. and Delic.V.  
TITLE Isolation and characterization of a cryptic plasmid from *Erwinia*

Pred. No. is the number of results predicted by chance to have a

JOURNAL J. Appl. Microbiol. 83 (4), 485-492 (1997)  
MEDLINE 98012514  
PUBMED 9351229  
REFERENCE 2 (bases 1 to 3661)  
AUTHORS Bilic Nezic, M. and Delic, V.  
TITLE Sequence Analysis and Functional Characterization of Replication Region in Plasmid pPZG500 from Pantoea citrea ATCC 31623  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 3661)  
AUTHORS Bilic Nezic, M. and Delic, V.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1999) Biotechnology, PULIVA d.d., Research Institute, Ptilaz baruna Filipovica 25, Zagreb 10 000, Croatia  
FEATURES  
source 1. 3661  
/organism="Pantoea citrea"  
/strain="ATCC31623"  
/db\_xref="ATCC:31623"  
/db\_xref="taxon:5336"  
/plasmid="pPZG500"  
complement(join(3188..3661,1..177))  
/note="RNAI"  
99..317  
/note="Orf1"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAD21205.1"  
/db\_xref="GI:4457230"  
/translation="MGQAQLCWGSNHPAYSIGALAKLKHKNLIGLGFLYOKKKIK  
QRKIFERNKILFEFRGDCYRFAIFKT"  
177  
/note="Orf1V"  
complement(334..342)  
/note="putative"  
/bound\_moiety="Dnaa"  
458..472  
/bound\_moiety="IHF"  
489..704  
/note="Orf2"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAD21206.1"  
/db\_xref="GI:4457231"  
/translation="MDCVKKALENEIEHYKKDGIKSFLOLHYFDIDKALNGDEG  
DIINSLSIDSLIFLMISOSAGLSIK"  
complement(845..1066)  
/note="Orf3"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAD21207.1"  
/db\_xref="GI:4457232"  
/translation="MPSGSGSIGSYDIFNFSCKSSSINFSSTLKHRSQIILRAD  
FLYLLEFRNOYLILFYISYGFISCVII"  
complement(1193..1348)  
/note="Orf4"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAD21208.1"  
/db\_xref="GI:4457233"  
/translation="MODLFCWHLFELHPIFYTHHFLFLYLKFSMMKFLSTERAENI  
SHIFLKK"  
1991..2365  
/note="Orf5"  
/codon\_start=1  
/transl\_table=1  
/product="putative heat-shock protein"  
/protein\_id="AAD21209.1"  
/db\_xref="GI:4457234"

CDS  
/translation="MDEKSRKLTLEKKDLKLEYASESLAKKLGTOGVLENTMKE  
MAKNINIDVNEETAGQYRKLFKNKVEHSKSDLYVGLDECGRNSFDKTRAFRCIC  
ERIQWIDIEFTTKLIQYKEKO"  
2362..2991  
/function="putative resolution of catenane plasmid dimers  
prior to plasmid segregation to the daughter cells"  
/note="Orf6"  
/codon\_start=1  
/transl\_table=1  
/product="lambda-integrase like protein"  
/protein\_id="AAD21210.1"  
/db\_xref="GI:4457235"  
/translation="WKEADYMODLIFKRLKNNKSYDRYAGFLSTCSTIGCPRAEYLK  
IEIVRNREDEGISFKLILGAKVNDGQSERLHFDLSKHNDENMNTLSQDKNKEF  
YKPDGKLYNSLROYLYIQHRTFSLYTLNHRASDLKASGADDFITIAFLGHRVTSOE  
FYGARSSKGGICIAVTGVCSDVVANKSQFAVSRPSQISTLSKHFKE"  
3154..3159  
-35\_signal 3175..3180  
-10\_signal 3188..3288  
misc\_RNA /note="RNAI"  
-10\_signal 3295..3300  
-35\_signal 3311..3316  
BASE COUNT 1366 a 472 c 730 g 1093 t  
ORIGIN  
Query Match 57.7%; Score 1066.6; DB 1; Length 3661;  
Best Local Similarity 98.9%; Pred. No. 6.3e-157;  
Matches 1128; Conservative 0; Mismatches 4; Indels 9; Gaps 5;  
QY 710 AAGCTTGCGCTTCGTGGGTGCAAGCGCTCAATGCTGATCGTGTGCTACTGTATA 769  
DB 1 AAGCTTGCGCTTCGTGGGTGCAAGCGCTCAATGCTGATCGTGTGCTACTGTATA 60  
QY 770 ACTTTCACCTTTTTCAGAGTAATTCAGAGGCGAAACATAGGTCACGCCAGCTTTCCT 829  
DB 61 ACTTTCACCTTTTTCAGAGTAATTCAGAGGCGAAACATAGGTCACGCCAGCTTTCCT 120  
QY 830 GGGGTTCGGCACATCCAGCTTACAGACATGTGTGCTCTTTCGACAGTCGAACAAATTC 889  
DB 121 GGGGTTCGGCACATCCAGCTTACAGACATGTGTGCTCTTTCGACAGTCGAACAAATTC 180  
QY 890 TAAATCCAGGTTTGGGTTTTTATACGAGAGCAAAACAAAATAAACAAGAAAAA 949  
DB 181 TAAATCCAGGTTTGGGTTTTTATACGAGAGCAAAACAAAATAAACAAGAAAAA 240  
QY 950 TTTTCGACGCAAAAAATATTTTGAATTTTAAAGCGATCTTGCTACCGCACTTTTG 1009  
DB 241 TTTTCGACGCAAAAAATATTTTGAATTTTAAAGCGATCTTGCTACCGCACTTTTG 300  
QY 1010 CCATATTTAAACCTGACTATCTTATATAGTTAA--TAGATATATCCGTTAGATTATA 1066  
DB 301 CCATATTTAAACCTGACTATCTTATATAGTTAA--TAGATATATCCGTTAGATTATA 360  
QY 1067 AGTATGTTAAAGAGATTAACAAATCACTATATATTTAATCTGTAATATATTTGAC 1126  
DB 361 AGTATGTTAAAGAGATTAACAAATCACTATATATTTAATCTGTAATATATTTGAC 420  
QY 1127 AGTGATTTAATATATTAAGAGATATATCTATTAGCTTAATATATACTAAAAAGAG 1186  
DB 421 AGTGATTTAATATATTAAGAGATATATCTATTAGCTTAATATATACTAAAAAGAG 480  
QY 1187 GTAAATATATGATTTGTATTTAAAAAGCATTAAGAAATGAATAGAACATTTATAA 1246  
DB 481 GTAAATATATGATTTGTATTTAAAAAGCATTAAGAAATGAATAGAACATTTATAA 540  
QY 1247 AAGAGCGGATATCAAACTCTTCTTACAATCTTGCAATCTTACTTGTATAGATAAGCAT 1306  
DB 541 AAGAGCGGATATCAAACTCTTCTTACAATCTTGCAATCTTACTTGTATAGATAAGCAT 600  
QY 1307 TAAATGATGATATGTCGATATTAATTAACCTCAAAATTTATCCATGATGAAGTTTG 1366  
DB 601 TAAATGATGATATGTCGATATTAATTAACCTCAAAATTTATCCATGATGAAG--TTTG 659

QY	1367	ATCTCTTGATGCTGTAGACACAATTTGGCTGGGCTTCAATAAAAATATACAGAGACA	1426
Db	660	ATCTCTTGATGCTGTAGACCAA---TCGGCTGGCTTCAATAAAAATATACAGAGACA	716
QY	1427	ATGAATATTTATCATCAGCTAAACCTGAAAAATGATTTTAAAAAATACCTGTTCTTATTC	1486
Db	717	ATGAATATTTATCATCAGCTAAACCTGAAAAATGATTTTAAAAAATACCTGTTCTTATTC	776
QY	1487	ATTGATCAATTTGGGAGAGATTTATTTACGATGAGATGATGATCAATCATCAACAAATGA	1546
Db	777	ATTTCATCATATTTGGGAGAGATTTATTTACGATGAGATGATGATCAATCATCAACAAATGA	836
QY	1547	TTAAAGATTTAGATATTTACACATTTGAGAGAAATTAACATATGATATAATAAATTAAGAGA	1606
Db	837	TTAAAGATTTAGATATTTACACATTTGAGAGAAATTAACATATGATATAATAAATTAAGAGA	896
QY	1607	ATATATTTGATTTGGAATTAACAAAAGATATTAAGAAGATGACGCTTAAAAAATATTCGC	1666
Db	897	ATATATTTGATTTGGAATTAACAAAAGATATTAAGAAGATGACGCTTAAAAAATATTCGC	956
QY	1667	ATTGATCGATGTTTTTAAAAAGTAGAGAAAAATTTATTCATGACAGAGATTTTGCACTAAA	1726
Db	957	ATTGATCGATGTTTTT---AAAGTAGAGAAAAATTTATTCATGACAGAGATTTTGCACTAAA	1015
QY	1727	GATATTTGAAAAATATTCGTATGAAAAATTCCTATTGATGATCCAGATGATGCGCATTAAGAAATC	1786
Db	1016	GATATTTGAAAAAATATTCGTATGAAAAATTCCTATTGATGATCCAGATGATGCGCATTAAGAAATC	1075
QY	1787	ACAATTTCCGACGAGAAAAATGCGCTATGCTTTTCGCTTTAAAAAAACCAACAAAAAGAGATC	1846
Db	1076	ACAATTTCCGACGAGAAAAATGCGCTAT---CTTTCAGCTTTAAAAAAACCAACAAAAAGAGATC	1134
QY	1847	T 1847	
Db	1135	T 1135	

	RESULT 2			
: AX345915/C				
LOCUS	AX345915	7597 bp	DNA	linear PAT 01-FEB-2002
DEFINITION	Sequence 986 from Patent WO0200928.			
ACCESSION	AX345915			
VERSION	AX345915.1 GI:18493801			
KEYWORDS	.			
SOURCE	synthetic construct. artificial sequences.			
ORGANISM	1			
REFERENCE	Olek,A., Piepenbrock,C. and Berlin,K. Diagnosis of diseases associated with the immune system Patent: WO 0200928-A 986 Q-JUN-2002; Epigenomics AG (DE)			
AUTHORS				
TITLE				
JOURNAL				
FEATURES	Location/Qualifiers			
source	1..7597 /organism="Synthetic Construct" /db_xref="taxon:32630" /note="Chemically treated genomic DNA (Homo sapiens)"			
BASE COUNT	1965 A 109 C 1790 G 373 T			
ORIGIN				
Query Match	7.0%; Score 130; DB 6; Length 7597;			
Best local similarity	47.7%; Pred. NO. 3.7e-11;			
Matches 447; Conservative	0; Mismatches 485; Indels 6; Gaps 2;			
OY	912 ATCCAGAAAGCAAAAATAAATTAACAAGAATAATTTTCGAGCAAAAAATTATTTT 971			
Db	1062 ATAATTAAATTAATTAATTAATTAATTAATTAATTAATGAATCATAAATCATATAA 10030			
OY	972 GGATTTTTTAAAGCGACTCTGTACCAGCACTTTGCCATATTTAACCTGCATATC 10311			
Db	1002 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 943			
OY	1032 TTATTAAGTAAATAGATATATCCGTTAGATATATAAAGTAGTTAAAAAGAGTA AAA CA 10911			

[illegible]

RESULT 3				
AX251345/c				
LOCUS	AX251345	6286 bp	DNA	linear
DEFINITION	Sequence	313 from Patent WO0168912.		PAT 05-OCT-2001
ACCESSION	AX251345			
VERSION	AX251345.1	GI:15984768		
KEYWORDS	.			
SOURCE	synthetic construct.			
ORGANISM	synthetic construct.			
REFERENCE	artificial sequences.			
AUTHORS	1 (bases 1 to 6286)			
TITLE	Olek,A., Piepenbrock,C. and Berlin,K.			
JOURNAL	Diagnosis of diseases associated with tumor suppressor genes and oncogenes			
FEATURES	Patent: WO 0168912-A 313 20-SEP-2001.			
	Epigenomics AG (DE)			
	Location/Qualifiers			

[illegible][illegible]





Overlapping clones are noted at the beginning and end of the features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards. Estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT:

Location/Qualifiers

source

1. .111861  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-384L1"  
1. .2005  
/note="overlaps bases 165307..167311 of clone AC073895"  
/function="clone overlap"  
complement(1195..1708)  
repeat\_region  
/rpt\_family="LTR33"  
1882..2418  
/rpt\_family="HAI1"  
2563..2586  
/rpt\_family="(TG)n"  
2618..2692  
/rpt\_family="(CATANA)n"  
2935..3049  
/rpt\_family="MIR"  
complement(3700..3783)  
repeat\_region  
/rpt\_family="MIR"  
4827..4848  
/rpt\_family="AT-rich"  
4981..5016  
/rpt\_family="AT-rich"  
complement(5037..5469)  
repeat\_region  
/rpt\_family="L1"  
complement(5721..6053)  
repeat\_region  
/rpt\_family="L1"  
6185..6231  
/rpt\_family="AT-rich"  
7100..7297  
/rpt\_family="MIR"  
8934..9209  
/rpt\_family="MER102"  
9667..9968  
/rpt\_family="AluSx"  
complement(10412..10660)  
repeat\_region  
/rpt\_family="LTR33"  
10710..10769  
/rpt\_family="MIR"  
10941..11011

repeat\_region  
/rpt\_family="(TA)n"  
11092..11196  
/rpt\_family="T-rich"  
complement(11197..11506)  
repeat\_region  
/rpt\_family="AluSx"  
11507..11599  
/rpt\_family="T-rich"  
12304..12379  
/rpt\_family="MER21B"  
12405..13097  
/rpt\_family="MER21B"  
13104..13257  
/rpt\_family="MLT1B"  
13304..13346  
/rpt\_family="(TA)n"  
13849..14101  
/rpt\_family="LIME2"  
14325..14431  
/rpt\_family="LRR16B"  
14446..14536  
/rpt\_family="LIME2"  
complement(15499..15556)  
repeat\_region  
/rpt\_family="MIR"  
complement(15509..15626)  
/rpt\_family="L2"  
complement(15700..15761)  
/rpt\_family="MER92C"  
16137..16200  
/rpt\_family="(TA)n"  
17077..17117  
/rpt\_family="AT-rich"  
19849..19949  
/rpt\_family="GA-rich"  
20599..20717  
/rpt\_family="FLAMC"  
20721..20763  
/rpt\_family="(TAAA)n"  
21442..21725  
/rpt\_family="MLT1A1"  
complement(21762..22107)  
repeat\_region  
/rpt\_family="MLT1"  
22360..22395  
/rpt\_family="AT-rich"  
22559..22817  
/rpt\_family="AluSx"  
complement(22927..23015)  
repeat\_region  
/rpt\_family="L2"  
23327..23601  
/rpt\_family="AluJo"  
23602..23623  
/rpt\_family="(TAAA)n"  
complement(23985..24086)  
repeat\_region  
/rpt\_family="L2"

#### Query Match

Best Local Similarity 48.0%; Pred. No. 6.2e-09; Length 111861;  
Matches 422; Conservative 0; Mismatches 448; Indels 9; Gaps 3;

QY 909 TTATACGAGCAAGCAAAACAAATAAACAAGAAATTTTCGAGCGCAAAATAT 968  
DB 43416 TAT 43475  
QY 969 TTTGGAATTTTAAAGCGATCTGCTACCGCACTTTGCCATTTTAAACCTGACT 1028  
DB 43476 TAT 43535  
QY 1029 ATCTTATATAGTTAATA-GATATATCGTTAGATTATTAAGTAGTAAAAAGATAA 1087  
DB 43536 TAAAT 43595  
QY 1088 AACATTAATCTTAT 1146  
DB 43596 AAT 43655

OY		1147	AGACATATATCTATTAGCTTTAAATCTTAACTAAAAAAGAGTAATATATGCGATTGGCTA	1206
Dd		43656	AAAATRTATATATATCTACTAATATATATAATATATATATATATATATATATATATATTA	4371
OY		1207	TTTTAAAAAGCATTAGAAAAATGAATAGAACATTATATAAAAAGAAGCGTGATCAATCT	1266
Dd		43716	ATA	4377
OY		1287	TTCCTACAATCTGCATTACTTGATATATACATAAACCATTAAATGGGTGATGATGCC	1326
Dd		43776	AAT	4383
OY		1327	GATATTTAAACCTCAAAATTTATCCATGATGAAAGTTTGATCTCTTGATGTGAGCAC	1386
Dd		43836	AAAT	4389
OY		1387	AATTTGGCGTGGCTTCATATAAAAATATACAGAGACGAATGAATTTTATCTCAAGCT	1446
Dd		43896	ATAAATA	4395
OY		1447	AAACCTGAAATGATTTTAAAAAATCTCGTCTCTTATTCATCTGCATCAATTGGAGAA	1506
Dd		43956	AAAT	4400
OY		1507	TTTATATTCGATGAGTGCATACATACATCAGAAATGATTAAAGCATTTAGATTAATTC	1566
Dd		44009	TAT	4406
OY		1567	ACATATGAGAGAAATACCATATGATATATATATATATATATATATATATATATATATAT	1626
Dd		44069	AT	4412
OY		1627	CAAAAGAT	1686
Dd		44129	ATA	4418
OY		1687	TGATATGAAAAATTTATATGATGAAGAAGATTGCAACTTAAGATATATGAAAAATATCGATG	1746
Dd		44189	TAAAT	4424
OY		1747	AAAAATCTATTTGATGATCCAGATGATGGCATATAGAAAT	1785
Dd		44249	TAT	44287
RESULT 7				
AX251284/c				
LOCUS	AX251284	7814 bp	DNA	linear PAT 05-OCT-2001
DEFINITION	Sequence 252 from Patent WO0168912.			
ACCESSION	AX251284			
VERSION	AX251284.1 GI:15984707			
KEYWORDS				
SOURCE				
ORGANISM	synthetic construct. artificial sequences. 1 (bases 1 to 7814)			
REFERENCE	Olek,A., Piepenbrock,C. and Berlin,K.			
AUTHORS	Diagnosis of diseases associated with tumor suppressor genes and oncogenes			
FILE	Patent: WO 0168912-A 252 20-SEP-2001; Epigenomics AG (DE)			
JOURNAL				
FEATURES				
source	Location/Qualifiers 1..7814			
	/organism="synthetic construct" /db_xref="taxon:32630"			
	/note="Chemically treated genomic DNA (Homo sapiens)"			
BASE COUNT	1677 a 101 c 1779 g 4257 t			
ORIGIN				

[illegible]



REFERENCE 2 (bases 1 to 190801)  
 AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxler, E., Blazek, R.G., Brehm, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Keaney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomonte, M.A., Macada, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Snir, E., Sylvas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zierian, L.L. and Rudin, G.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 COMMENT On Aug 24, 2001 this sequence version replaced gi:6957960. Sequence submitted by: Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).  
 FEATURES  
 source 1. 190801  
 /organism="Drosophila melanogaster"  
 /strain="y. cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 /map="82E-82F"  
 /clone="BACR01A18 (D669)"  
 /clone\_11b="Rpci-98 (Rosewell) Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in PBACs3.6")  
 BASE COUNT 53774 a 39988 c 41097 g 55942 t  
 ORIGIN  
 Query Match 6.1%; Score 112.6; DB 3; Length 190801;  
 Best Local Similarity 48.2%; Pred. No. 9.9e-09;  
 Matches 407; Conservative 0; Mismatches 434; Indels 4; Gaps 3;  
 Oy 910 TTATACGAAGCAAAACAAATAAATAACAGAAATTTTGCAGCAAAATATAT 969  
 Db 180257 TCATATAAAATTAATAATAATAATAATAATAATAATAATAATAATAA 180198  
 Oy 970 TTGCAATTTTAAAGCGATACCTGACCGACCTTTGCCATATTTAAACCTG 1029  
 Db 180197 AAGTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 180138  
 Oy 1030 TCTTTAATAGTAATAGATATATCGTTAGATTAATAAGTATGTTAAACGAGTAA 1089  
 Db 180137 AATGAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 180078  
 Oy 1090 CAAT--AACTATATATTTAATTCGAATTAATTTGACAGTATTTAATATTA 1147  
 Db 180077 AAAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 180018  
 Oy 1148 GAGATATATCTATTAGTTAATAATTAATAATAATAATAATAATAATAATA 1207  
 Db 180017 AAAATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179958  
 Oy 1208 TTAATAAAGCATTAAGAAATGAATAGAAATTAATAATAAAGCGGTATCAATCT 1267  
 Db 179957 TAGTAAATAGTAAATAGTAAATTAATAATAATAATAATAATAATAATA 179899  
 Oy 1268 TCTTACATACCTGATTAATTTGATAGATAGAAAGCTTAATGATGATGCGG 1327  
 Db 179898 AAAATATAATAATAATAAGAAACATTAATAATAATAATAATAATAATAAT 179839  
 Oy 1338 ATATTATAACTCAATTTATTCATGATGAAGAGTTTGATCTCTTGATGAGACA 1387  
 Db 179838 TAATGATTAATTCATTAATAATAGGAATTAATAATAATAATAATAATAATTA 179779

Oy 1388 ATTTGCGTGGGCTTCAATAATAATATACAGACGAATGAATTTTATCATCAGCTA 1447  
 Db 179778 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179719  
 Oy 1448 AAACGAAATGATTTTAAATAATACCTGCTTTATTCATTCGATTCATTCGAGAA 1507  
 Db 179718 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179659  
 Oy 1508 TTAATTCAGATGAGATGATGATCAATACATCAAGAAATGATTAAGATTAATTA 1567  
 Db 179658 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179599  
 Oy 1568 CATATGGAATTAACCATGATGAATTAATAATAATAAGAAATATATGATTCGAATAC 1627  
 Db 179598 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179539  
 Oy 1628 AAAAGATTAATAAAGAACTCAGCTCTTAAATAATCTCATGATGATGTTTAA- 1686  
 Db 179538 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179479  
 Oy 1687 TAGATGAATAATTTTATGATGAAGAGATTGCAACTTAAGATATTCGAATATCGTATG 1746  
 Db 179478 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179419  
 Oy 1747 AAAAT 1751  
 Db 179418 TAAAT 179414

RESULT 10  
 AX347260/c 12237 bp DNA linear PAT 01-FEB-2002  
 LOCUS AX347260  
 DEFINITION Sequence 2331 from Patent WO0200928.  
 ACCESSION AX347260  
 VERSION AX347260.1 GI:18495148  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM artificial sequences.  
 REFERENCE 1  
 AUTHORS Olek, A., Piepenbrock, C., and Berlin, K.  
 TITLE Diagnosis of diseases associated with the immune system  
 JOURNAL Patent: WO 0200928-A 2331 03-JAN-2002;  
 Epigenomics AG (DE)  
 FEATURES  
 source 1. 12237  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="Chemically treated genomic DNA (Homo sapiens)"  
 BASE COUNT 3050 a 207 c 3024 g 5956 t  
 ORIGIN

Query Match 6.0%; Score 111.6; DB 6; Length 12237;  
 Best Local Similarity 48.5%; Pred. No. 2.5e-08;  
 Matches 364; Conservative 0; Mismatches 384; Indels 2; Gaps 2;  
 Oy 1014 ATTAAAACTGACTATCTTATTAAGTTAATAGATATATCCGTAGATTAAGTATGT 1073  
 Db 6912 AATCAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6853  
 Oy 1074 TAAAAACGATAAAACATATATATATTTAATTCGATTTATTTGACAGTAT 1133  
 Db 6852 AATCAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 6793  
 Oy 1134 ATTATATATTAAGAGATATATCTATAGCTTA-ATATAACTAAAAAGAGCTAAT 1192  
 Db 6792 AATAATTAATCAATAATAATAATAATAATAATAATAATAATAATAATAAT 6733  
 Oy 1193 ATATGATGTTGTTATTTAAATAAAGCATTAAGAAATGAATATGAACATTATTA 1252  
 Db 6732 AAATAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6674

```

QY 1253 GTGATATCAATCTTCTTACAAATCTTCATTCCTTGATATAGATTAAGCATTAATG 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6673 TAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1313 GTGATGATGTGGGATATATTAACCAATTTATTCATTCATGATGAAGCTTTGCTTC 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6613 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1373 TTGATGTTGACCAATTTTCGGCTGGCTTTCAATTAATAATTAATTAATTAATTAAT 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6553 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1433 ATTATATCATGCTGCTAAACCTGAAATGATTTTAAATAATCTGCTCTTTATTCATTCGA 1492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6493 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1493 TCAATTTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6433 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1553 GATTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6373 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1613 TGATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6313 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1673 CGATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6253 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1733 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6193 AATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
AE003603/c 294914 bp DNA linear INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386043 section 5
ACCESSION AE003603
VERSION AE003603.2 GI:10727164
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 294914)
          Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
          Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
          George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
          Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
          Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
          Miklos,G.L., Adair,J.F., Agbayani,A., An,H.J.,
          Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
          Binkendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
          Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
          Botchan,M.R., Bouck,J., Brokstein,P., Brottler,P., Burtis,K.C.,
          Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
          Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
          Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
          Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
          Dunn,P., Durbin,K.J., Evangelista,C.C., Ferrariz,C., Ferrier,S.,
          Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
          Gelibert,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
          Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
          Hernandez,J.R., Houck,J., Hostlin,D., Houston,K.A., Howland,T.J.,
          Wei,M.H., Ibegwam,C., Jaisl,M., Kalush,F., Karpen,G.H., Ke,Z.,
          Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
          Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
          Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Ma,J., Maizels,B., McIntosh,T.C.,
          McLeod,M.P., McPherson,D., Mekulov,G., Mitsuhashi,N.V., Mobarry,C.,
          Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
          Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
          Nusskern,D.R., Pacle,J.M., Palazolo,M., Pitaru,G.S., Pan,S.,
          Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
          Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
          Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
          Stapleton,M., Strong,R., Sun,E., Svitek,R., Tector,C., Turner,R.,
          Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
          Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
          Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
          Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
          Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
          Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
          The genome sequence of Drosophila melanogaster
          Science 287 (5461), 2185-2195 (2000)
TITLE JOURNAL
MEDLINE 20196006
PUBMED 10731132
REFERENCE 2 (bases 1 to 294914)
AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
JOURNAL Direct Submission
TITLE Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
          Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gi:7296751.
FEATURES
          location/Qualifiers
          1..294914
          /organism="Drosophila melanogaster"
          /db_xref="taxon:7227"
          /chromosome="3R"
          complement(<8081..>8296)
          /gene="CG14663"
          /evidence=not_experimental
          /map="82F1-82F1"
          /db_xref="FLYBASE:FBan0014663"
          /db_xref="FLYBASE:FBgn0040683"
          complement(<8081..>8296)
          /gene="CG14663"
          /evidence=not_experimental
          /product="CT34443"
          /db_xref="FLYBASE:FBan0014663"
          /evidence=not_experimental
          /db_xref="FLYBASE:FBgn0040683"
          complement(8081..8296)
          /note="CG14663 gene product"
          /codon_start=1
          /evidence=not_experimental
          /protein_id="AAFS2070.1"
          /db_xref="GI:7296792"
          /db_xref="FLYBASE:FBan0014663"
          /db_xref="FLYBASE:FBgn0040683"
          /translaton="MPRKLPAARNASMLVPCLLPLPARMPARPPGVPPAPRSICR
          ODIPLEGICNMSHIGETELVMD"
          <13958..>14959
          /gene="CG12590"
          /evidence=not_experimental
          /map="82F1-82F1"
          /db_xref="FLYBASE:FBan0012590"
          /db_xref="FLYBASE:FBgn0037294"
          <13958..>14959
          /gene="CG12590"
          /product="CT34442"
          /evidence=not_experimental
          /db_xref="FLYBASE:FBan0012590"
          /db_xref="FLYBASE:FBgn0037294"
          13958..14959
          /gene="CG12590"
          /note="CG12590 gene product"
          /codon_start=1
          /evidence=not_experimental
          /protein_id="AAFS2069.1"
          /db_xref="GI:7296791"

```



QY 1639 AAGAAAGTACGCTTTAAATATCGATTCGATGCTTTTAAAGATGAAAT 1698  
DB 3952 ATAAATTAATAAGTAAATTAATAATTAATAACCAAAAAATGTTGATATT 3893  
QY 1699 TATATGATGAGAGATTTGCACTAAGATATGAAATATCGAATGAAATCCATATG 1758  
DB 3892 AATGATACGATACATTTAATACATTTACATTTAATGATTTTATGAAACCAATGC 3833  
QY 1759 ATG 1761  
DB 3832 TTG 3830  
RESULT 12  
PFMALP3 67970 bp DNA linear INV 15-DEC-1999  
LOCUS Plasmodium falciparum MALP3, complete sequence.  
ACCESSION AL031746  
VERSION AL031746.9 GI:6594243  
KEYWORDS HTG.  
SOURCE Plasmodium falciparum 3D7.  
ORGANISM Plasmodium falciparum 3D7.  
REFERENCE 1 (bases 1 to 67970)  
AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quall, M., and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
-COMMENT  
On Dec 16, 1999 this sequence version replaced gi:5763807.  
For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/projects/P\\_falciparum](http://www.sanger.ac.uk/projects/P_falciparum). IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.  
FEATURES  
source 1..67970  
Location/Qualifiers  
/organism="Plasmodium falciparum 3D7"  
/strain="3D7"  
/db\_xref="taxon:36329"  
/chromosome="1"  
/complement(join(1748..2598,2748..2848,2990..3276))  
/gene="MALP3.01"  
/complement(join(1748..2598,2748..2848,2990..3276))  
/note="MALP3.01, conserved hypothetical protein, len: 412 aa, similarity: UPP0006 family eg to YBL055C/YBL0512/YBL0511, YBP5\_YEAST (418 aa), fasta scores: opt: 316, E( ): 1.1e-12, (33.2% identity in 271 aa overlap)"  
/codon\_start=1  
/product="conserved hypothetical protein, UPP0006 family"  
/protein\_id="CAB63556.1"  
/db\_xref="GI:6594244"  
/translation="MKLVHYIKYINLVFISIIPLKSNLSIKYNDLRYSIVNKKYK LOIKRSNLKKNNHNRKEDNESSFDIGSNLTKMPGVNNSKKHENDLONVBRK NNNYDKIITCTCAIEDKSLKICTYDPECKFLYLSGVPNTCYETIDKRNHEKE ITAKKEYEFTIKRNEOVNSKMGKICDGGKNNNLNELLERNLDTIPGEKY NEKDEYELNKKIKYKPNRIVCGEIGLDPRLYFCSYIQIKYIFPOLQVOMN LPPLHMRNCSETFEKIVDIYKFLPEKNGGVHSTFDKEDIHIIIVCNKLYGVNG CSLKLEINAVKRIPLNLLETDAPMGCVKTHASVEYIKDTYEKRAYNLKKIKN IIKDDNITFEKNEPNYIA"  
misc\_feature complement(2599..2610)  
/gene="MALP3.01"  
/note="potential splice acceptor sequence"  
misc\_feature complement(2742..2747)  
/gene="MALP3.01"  
/note="potential splice donor sequence, atg/gttaa"  
misc\_feature complement(2849..2861)

misc\_feature  
gene  
CDS  
misc\_feature  
gene  
CDS  
/gene="MALP3.01"  
/note="potential splice acceptor sequence"  
/complement(2984..2989)  
/gene="MALP3.01"  
/note="potential splice donor sequence, aaa/gtaaaa"  
5005..5496  
/gene="MALP3.02"  
5005..5496  
/gene="MALP3.02"  
/note="MALP3.02, hypothetical protein, len: 163 aa, contains possible signal sequence"  
/codon\_start=1  
/product="hypothetical protein, MALP3.02"  
/protein\_id="CAB63557.1"  
/db\_xref="GI:6594245"  
/translation="MKLNNRFVAVCPILIIFFFLNSVVLGNNNNNINPHETENAAK AMRKLGEINSIKLDNGDELKIKDKERKHDSPWQKSYSPISLSEEXYSQDPLR KKEINENANTKIIDROEFYILNNDIEINATRFVLENNDELIIQSKSLIDIIOS LNN"  
8020..10389  
/note="possible cen1, region of very high [A+T] content"  
1484..20352  
/gene="MALP3.03"  
1484..20352  
/gene="MALP3.03"  
/note="MALP3.03, putative ABC transporter, len: 1822 aa"  
/codon\_start=1  
/product="putative ABC transporter"  
/protein\_id="CAB63558.1"  
/db\_xref="GI:6594246"  
/translation="WTYKENVGISNKKKKKSCONISDFLNSFDMIRPLINDLIK GDIDLPNCRNFDVPYVASKLENDLIDVEDSEFSYKSSNMFVLAHNSDASE KVVYVYVYHNILMSLKTKFERIILISFYLETILVIGKFDIYVRILIEGQKIPV YISFLKDFKVSGLVVMIMFEPFLFEALHFEYFLINLKSVMFLYLYKICSN NHLOPDAFYMTYRKESQTEIDEISDFSLISGKMASSSSGKIKNNKINDNKEVN DYIINPESKTKMEQDSNENRSDNPNVYIMDSVPEFPTSCINLTPVYKIF MSFYFHRTKISNSGVAIWSIALYSMAIIPFLPSFKSKYLIRKRDNNHVL KEFLIKRNMSEFAKTYINIFRMKEMVCKIRYLSNIGVFISISDIYEVEYFEL YLKRNLNKEEIKFTSIIMPLYVVKILISNVAENPNLVNVMGIVNTRLNLYINDH LYNDIKNYFMYRTYRNDYNIYVDKTLQDENTISHDGTSNHLKHKYNTKRLTN MEKYFEFFKNNYHNKNIINKOILSGLKNVDNNTYKICPOEHSNSYVNNSSHIE KKEEYENHINSNSGSMNEFEKKKNNYIITKLENGSGLSYDNKCONDHLIKNNEN LKRNSLIITIGVSGKSAFPHSILIGDLMNTHGUILYENFPKRPILYIPONSMLPG NTRSMITLGENEPLIYKYTLQSELDNDLSTIEGDGKYLINDHNPISGKQVRCLA BALVEHYIHMHRKLDYDKKLIQPMELIKOLINNKINSYNNKSKLVANIPENEN YLOKCMDNNFYLYLDIDITSLDPSISKFTFLPCKENISFKDNCSTLISNKS TLDFNLELIDLDNOYEVNIIEIDKTLKYAGNSSEYKNNLNTIKESHGYSMLNT IDYRIRIKLFDSEVLNHYVKSNNKMYKEAYFVKGNTESEYSPEDISNKEYIKKKKNY KKEHNKNNKNNNNNNNNKNDHINIMNNHRYNDINLGNPSTDSPPVYSLGNE YTLDTYISNSDKKEIYAPLYKDTHEEENKSSMPEFVASSNMTINPSENFEEDNSS FKSISLSTYLYMFOQGVFLVLTSVIEMLSIFDEIKFVFLMMSIISNNKESHD TILQKOVLYLYMFLYPLISLVTSGICSMIYGNITSFALVHNMIISINAKLYIF YNNNGIYNINFEIYDISAFDGEFLRIYKAFIEPRCLISLLIYMRDCLIFEPYV IILYFEPFERESGCKEAPORLYLSCHLNTIYISNLSGNNIINIYKNNYKNDVVE HYINNRISYEPFKULINWASLYIKIFILLTYYTIIMHPHIASGIIILKEKYVRI LSTIGYCTISFARLGVIIKFLDCTTHIEKMCVQQLKEEFAKSNKSNASNNENEL NVITTYQYKRENNENISDKISAIVKKNVNSIINSQSDSKRYKRGTFYVYVSK KKIPLVNGTYIYIDEEPSLKNINWALNMOGIGVYSGKSAKSTLISLILINISOG KITYEGRDIPIYNNKGEDSIIGILAAOSFVYNNNIRFIDIPYNNFTMDLIVHAKLN GINLGNKDLTYVYKODKSNYKRIQSYKINSONDTILLTNDICIRYLSVLYLNN RHAKRIIILIDIEPIPLNLSNHSDELNSLIRKASFNPIIINHPNNTVLIISHANT LSCDYLIVLKRGEITYRCSEYEDVKYQSELSHLEMD"  
/gene="rRNA"  
23896..31533  
/gene="rRNA"  
23896..31533  
/note="region containing small subunit, 5.8S and large subunit rRNA genes and spacer regions"  
complement(join(31966..32476,32675..32775))  
/gene="MALP3.04"  
complement(join(31966..32476,32675..32775))  
/note="MALP3.04, conserved hypothetical membrane protein,







## COMMENT

Cambridgeshire, CB10 1SA, UK: E-mail enquiries: humbrey@sanger.ac.uk Clone request@sanger.ac.uk On Feb 1, 2002 this sequence version replaced g1:16973851.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, SW: SWISSPROT, Tr: TrEMBL, Wp: WormPEP; Information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Ch1 RPL1-269F19 is from the library RPL1-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6  
IMPORTANT: This sequence is not the entire insert of clone RPL1-269F19 it may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true left end of clone RP5-88207 is at 88488 in this sequence. The true right end of clone RP4-678E16 is at 2000 in this sequence.

## FEATURES

## source

1. 90487  
/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/clone="RPL1-269F19"

/clone\_1lb="RPL1-11.1"

2097. 2223

/note="Sequence from reads from a short insert library derived from a clone PCR. Restriction digest data confirm the assembly."

BASE COUNT 22057 a 22328 c 23459 g 22643 t  
ORIGIN

Query Match 5.9%; Score 108.8; DB 9; Length 90487;  
Best local Similarity 48.4%; Pred. No. 4.5e-08;  
Matches 449; Conservative 0; Mismatches 472; Indels 7; Gaps 5;

QY 922 CAAACAAAAAATAAACAAGAAAAATTTGAGCGAAAAAATATTTGAATTTTT 981  
DB 58868 CAAAAACAAAAAATAAACAAGAAAAATTTGAGCGAAAAAATATTTGAATTTTT 981  
QY 982 AAAGGAGTACTTGTACCGCACTTTGCGATTTTAAACCGGACTATCTTATAAGT 1041  
DB 58928 TAAAGGCGCATTTAGGCGCAAACTGTGCCCCCTCCCAAAATTCATATTAATAATATA 58987  
QY 1042 AATAGATATATTCGTTAGATTTAATGATTTAAACAGATTAACAGATTAATATA 1101  
DB 58988 AATATATATTAATATTAATATTAATATTAATATTAATATTAATATTAATATA 59047  
QY 1102 TATTTAATTCGATTTATTTGACAGTGATTTAATATATTAATTAAGATATATCTAT 1161  
DB 59048 TATTAATATTAATATTAATATTAATATTAATATTAATATTAATATTAATATA 59107  
QY 1162 AGCTTAATATTAATTAACAAAAAGGTAATATATGATTTGTATTTAAAAAGCATTA 1221  
DB 59108 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59167  
QY 1222 GAAAAATGAATAGACATTAATAAAA-AGACGCTGATATCAAACTCTTCTTACAATACTT 1280

DB 59168 TATATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59227  
QY 1281 GCATTACTTGTATATGATTAACCATTAATTTGGTGAATGCGCAT-ATTATTAAC 1339  
DB 59228 TAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59287  
QY 1340 CAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59347  
DB 59288 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59347  
QY 1400 CTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 1459  
DB 59348 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59407  
QY 1460 ATTTAAAAAATCTGTTCTTTATTCATTCGTCATTTGGGAAATTTAATAGATG 1519  
DB 59408 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59467  
QY 1520 AGATGATGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATTAATATA 1579  
DB 59468 ATATTAATA-TATACATATATACATATATTAATTAATTAATTAATTAATTAATATA 59526  
QY 1580 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 1639  
DB 59527 TATGATATAT-ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59585  
QY 1640 AGAAGTACCTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 1699  
DB 59586 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59645  
QY 1700 TATTTGATGAAGGATTTGCACTTAAGATTAATTAATTAATTAATTAATTAATATA 1759  
DB 59646 TATTAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59702  
QY 1760 TGATCCAGATGATGATTAAGAAAAATCAATTCGACGAGAAATGCTTTCG 1819  
DB 59703 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59762  
QY 1820 CATTAAAAAACAAGCAAAAAGAGTCT 1847  
DB 59763 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATATA 59790

RESULT 15  
AC103590 213056 bp DNA linear PRI 30-JAN-2002  
LOCUS  
DEFINITION Homo sapiens chromosome 1 clone RPL1-54K18, complete sequence.  
ACCESSION AC103590 AL357933  
VERSION AC103590.2 GI:18425292  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 213056)  
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.  
Direct Submission  
Unpublished  
2 (bases 1 to 213056)  
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
Direct Submission  
Submitted (29-NOV-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 213056)  
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.  
Direct Submission  
Submitted (30-JAN-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
On Jan 30, 2002 this sequence version replaced g1:17149455.  
----- Genome Center

Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: uwgchgs@u.washington.edu  
 Drafting Center: SC

----- Project Information

Center project name: chr-1  
 Center clone name: RP11-544K18 (sc0176)

----- Summary Statistics

Sequencing vector: plasmid; 6% of reads  
 Sequencing vector: plasmid; 108752; 94% of reads  
 Chemistry: Dye-terminator; Big Dye; 33% of reads  
 Chemistry: Dye-terminator; Big Dye; 33% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 212910 bases at least Q40  
 Consensus quality: 212999 bases at least Q30  
 Consensus quality: 213045 bases at least Q20  
 Insert size: 212874; sum-of-contigs  
 Quality coverage: 11.3x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5': RP11-249G23 AL360008, 2000-bp overlap  
 3': RP11-568G11 AL445435, 55927-bp overlap

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

----- EcoRI

----- HindIII

----- BglII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	9333	2083	2207	4498	4445
6	<800	6382	6525	2067	2067
8596	8488	512	<800	4570	4720
6075	6254	449	<800	8205	8225
744	803	3942	3884	6233	6198
1451	1430	33	<800	46	<800
5074	4975	4226	4220	6178	6198

8472	8488	2573	2547	1093	1117
791	803	2206	2207	1370	1329
1630	1634	2392	2547	580	<800
976	947	4969	4925	1805	1789
2759	2807	4673	4652	479	<800
310	<800	1874	1855	19	<800
4237	4205	1946	1960	3931	4030
3832	3797	3486	3458	3165	3213
3874	3907	197	<800	2530	2554
1078	1066	1107	1067	3866	3836
1295	1254	3787	3747	944	944
3092	3047	4066	4005	3013	3006
6272	6254	6168	6150	2222	2188
7581	7483	1075	1067	2324	2374
377	<800	1831	1855	704	<800
2328	2392	910	895	3306	3213
1985	1975	1344	1286	1282	1329
4079	4075	193	<800	7936	7908
9368	9333	405	<800	5478	5466
1675	1634	9929	10079	1144	1117
9377	9333	10756	10538	20101	20463
1996	2119	9627	9680	8397	8591
34994	34910	607	<800	8571	8591
309	<800	403	<800	4082	4030
838	803	10148	10538	5725	5466
2150	2119	10455	10538	5930	5737
6918	7078	6913	6973	635	<800
1374	1327	1677	1648	5211	5124
1770	1769	5476	5444	5529	5737
2610	2635	1075	1067	1393	1329
2246	2297	1434	1394	7324	7355
140	<800	595	<800	5699	5737
2470	2539	1075	1067	7123	7086
6	<800	17612	17124	3985	4030
2081	1975	18631	18193	603	<800
31	<800	5953	5976	2550	2554
5702	5672	2204	2207	1887	1914



**THIS PAGE BLANK (USPTC)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 21:08:10 ; Search time 420.275 Seconds  
(without alignments)  
9896.940 Million cell updates/sec

Title: US-08-876-132-2

Perfect score: 1847  
Sequence: 1 ACATCTCAACACAGTTTAA...AACACAGCAAAAGAGATCT 1847

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: N.Geneseq-101002: \*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1847	100.0	1847	20	AAV74276
2	130	7.0	7597	24	ABL33013
3	123.2	6.7	6286	22	AA546591
4	114.2	6.2	6109	24	ABL32326
5	114.2	6.2	6109	24	AA561077
6	113.6	6.2	7814	22	AA546530
7	111.6	6.0	12237	24	ABL34358
8	98.2	5.3	20420	22	AAK73165
9	98.2	5.3	20420	24	ABK69933

C 10	97.8	5.3	13321	22	AA546421	Tumour suppressor
C 11	97	5.3	5413	22	AA546694	Tumour suppressor
C 12	96.4	5.2	19965	22	AAK73166	Human immune/haema
C 13	96.4	5.2	19965	22	ABK69932	Human secreted pro
C 14	96.2	5.2	6419	24	ABL32367	Human immune syste
C 15	91.4	4.9	6154	24	ABL70167	Chemically treated
C 16	91.4	4.9	6154	24	AA561112	Human gene regulat
C 17	91.4	4.9	6154	24	ABK31200	Signal transductio
C 18	91.2	4.9	6292	22	AA546735	Tumour suppressor
C 19	90.2	4.9	7442	22	AA546866	Tumour suppressor
C 20	89.6	4.9	6048	22	ABG67002	Human angiogenesis
C 21	89.6	4.9	6106	22	AA546429	Tumour suppressor
C 22	89.6	4.9	6106	22	ABK40031	Human chemically p
C 23	88.6	4.9	6106	24	ABL33472	Human immune syste
C 24	88.6	4.8	15548	24	ABL34155	Human immune syste
C 25	86.2	4.7	9760	24	ABL70197	Chemically treated
C 26	86.2	4.7	9760	24	AA561155	Human gene regulat
C 27	86.2	4.7	9760	24	ABK31242	Signal transductio
C 28	86	4.7	6175	24	ABL33307	Human immune syste
C 29	86	4.7	8310	20	AA229911	CDNA encoding a SC
C 30	85.4	4.6	7498	24	ABL32257	Human immune syste
C 31	84.6	4.6	10048	24	ABL70313	Chemically treated
C 32	84.6	4.6	10048	24	AA561251	Human gene regulat
C 33	84	4.5	6831	24	ABL33487	Human immune syste
C 34	83.6	4.5	1452	21	AAK70123	Plasmodium falcipa
C 35	81.6	4.4	5689	22	AA545384	Chemically pretrea
C 36	81.6	4.4	5689	22	AA546426	Tumour suppressor
C 37	81.6	4.4	5689	22	ABK28226	DNA transcription
C 38	81.2	4.4	12639	24	ABN80107	Human chemically m
C 39	81	4.4	8961	24	ABL49380	Human polynucleoti
C 40	81	4.4	8961	24	ABK28428	DNA transcription
C 41	81	4.4	9662	24	ABL32357	Human immune syste
C 42	80.8	4.4	5893	24	ABL32859	Human immune syste
C 43	80.8	4.4	50000	24	ABL55643	AmpeV genome fragm
C 44	80.6	4.4	83391	24	ABG67093	Human anglogenesis
C 45	80.4	4.4	9810	24	ABL32427	Human immune syste

#### ALIGNMENTS

RESULT 1	AAV74276	standard; DNA: 1847 BP.
ID	AAV74276:	
AC	AAV74276:	
XX		
XX		
DT	16-JUN-1999	(first entry)
XX		
DE	Pantoea citrea cryptic plasmid PS.	
XX		
KW	Cryptic plasmid; PS; higher temperature; growth; elimination;	
KW	mobilization; ds.	
XX		
OS	Pantoea citrea.	
XX		
PN	WO9859054-A1.	
XX		
PD	30-DEC-1998.	
XX		
PF	22-JUN-1998; 98WO-US12945.	
XX		
PR	23-JUN-1997; 97US-0876132.	
XX		
PA	(GENV ) GENENCOR INT INC.	
XX		
PI	Causesy SC, Fowler T;	
XX		
XX	WPI, 1999-105625/09.	
DR		
XX		
PT	Method for preparing an improved Enterobacteriaceae strain - useful	
PT	for improving bacterial fermentation strains of the family	
PT	Enterobacteriaceae	

OY	1	AGATCTCAACAGATTAAATCGCACTTCAGAAACTAAATAAATGAGGGCGGACCCGGCTC	60
Db	1	AGATCTCAACAGATTAAATCGCACTTCAGAAACTAAATAAATGAGGGCGGACCCGGCTC	60
OY	61	TTTTTTTGATGTTTGTAGTATATCTGTTAGAGTCACAGAAAAAGCATTTT	120
Db	61	TTTTTTTGATGTTTGTAGTATATCTGTTAGAGTCACAGAAAAAGCATTTT	120
OY	121	AGAGGGTTGACGAGGTTTTTTCAGCTATCCAGGGTTTTTGCGTTTTTGGGTTTGATC	180
Db	121	AGAGGGTTGACGAGGTTTTTTCAGCTATCCAGGGTTTTTGCGTTTTTGGGTTTGATC	180
OY	181	AGAAAGTCGTTCAAGATTATTTGACATTAAGACAGAGAAGTTTATTAACAGATCCAGATA	240
Db	181	AGAAAGTCGTTCAAGATTATTTGACATTAAGACAGAGAAGTTTATTAACAGATCCAGATA	240
OY	241	CGACAAAACACAGCTTTGCGAGCGCTTTGGAAGACTAAAGAACGGGAGACTTCTTTGA	300
Db	241	CGACAAAACACAGCTTTGCGAGCGCTTTGGAAGACTAAAGAACGGGAGACTTCTTTGA	300
OY	301	GTCCTGTATCAAGTTGGTCAGAACTCGATTACGATTTTGTAAAGTGAAGATCTAATCTACA	360
Db	301	GTCCTGTATCAAGTTGGTCAGAACTCGATTACGATTTTGTAAAGTGAAGATCTAATCTACA	360
OY	361	TTTTGCGAAGAAAGTCAAACTTACCTTGTAGTTACACTGAAAAATTTCCAGCCTTTTGA	420
Db	361	TTTTGCGAAGAAAGTCAAACTTACCTTGTAGTTACACTGAAAAATTTCCAGCCTTTTGA	420
OY	421	ATCCTTAAGCATATATTTTGTTTAAACCGATTTGTCCGGTGTGTTGGTGTGAGCCAT	480
Db	421	ATCCTTAAGCATATATTTTGTTTAAACCGATTTGTCCGGTGTGTTGGTGTGAGCCAT	480
OY	481	TGATCCGAGTGGTCAATATGTGATTTGTCCGCAACAGTGTATGTAGCTTAAACGGGGA	540
Db	481	TGATCCGAGTGGTCAATATGTGATTTGTCCGCAACAGTGTATGTAGCTTAAACGGGGA	540
OY	541	GTGCTACAAAAGACCATATCCGAAACGAGTGGCTTAAGTTTGGTTATCAACACAGGTAA	600
Db	541	GTGCTACAAAAGACCATATCCGAAACGAGTGGCTTAAGTTTGGTTATCAACACAGGTAA	600
OY	601	GCTATGAGAAAGCCAGCCATTAATGGGGTTAGTTGAAGCAAGTCTTCATATGGTGGGA	660
Db	601	GCTATGAGAAAGCCAGCCATTAATGGGGTTAGTTGAAGCAAGTCTTCATATGGTGGGA	660
OY	661	CACAAAGGGGTAGTAGAGGTGTCGCAACAGAAAGGTTGTATAGCTCTAAGCTTGTGCT	720
Db	661	CACAAAGGGGTAGTAGAGGTGTCGCAACAGAAAGGTTGTATAGCTCTAAGCTTGTGCT	720
OY	721	TCTGAGGGTCAAGGCTCAAGTGTGATCTGAGTGTGCTGTACACCGATTAATCTTCACTT	780
Db	721	TCTGAGGGTCAAGGCTCAAGTGTGATCTGAGTGTGCTGTACACCGATTAATCTTCACTT	780

QY	781	TTCCAGTAAATTTACGAGGCGGAAACATATGGGCTCAACCCAGCTTCTGGGGTTGGCA	840
Db	781	TTCCAGTAAATTTACGAGGCGGAAACATATGGGCTCAACCCAGCTTCTGGGGTTGGCA	840
QY	841	CATCCAGCTACAGCATTTGGTGTCTCTGCGAAGCTGAAGCACAATAATCTAATCCAGGT	900
Db	841	CATCCAGCTACAGCATTTGGTGTCTCTGCGAAGCTGAAGCACAATAATCTAATCCAGGT	900
QY	901	TTGGGTTTTTATACGAGACAAACAAAAATTAACACAAGAAAAATTTTGCACGA	960
Db	901	TTGGGTTTTTATACGAGACAAACAAAAATTAACACAAGAAAAATTTTGCACGA	960
QY	961	AAAAATATTTGGAAATTTTTTAAAGGCGATCTTGTCACGCACTTTGGCATTTTAAA	1020
Db	961	AAAAATATTTGGAAATTTTTTAAAGGCGATCTTGTCACGCACTTTGGCATTTTAAA	1020
QY	1021	ACCTGACTATCTTTATTAAGATTAATAGATATATCCGTTAGATTATTAAGTATGTTAAAC	1080
Db	1021	ACCTGACTATCTTTATTAAGATTAATAGATATATCCGTTAGATTATTAAGTATGTTAAAC	1080
QY	1081	GAGTAAACAAATACATATATTTAATCTGAAATATATTTGACAGTATATTAAT	1140
Db	1081	GAGTAAACAAATACATATATTTAATCTGAAATATATTTGACAGTATATTAAT	1140
QY	1141	ATATTAAAGATATATCTATTAGCTTAAATTAACATAAAGGATAATATATGAT	1200
Db	1141	ATATTAAAGATATATCTATTAGCTTAAATTAACATAAAGGATAATATATGAT	1200
QY	1201	TGTGATTTAAAAAGCATTTGAAATGAATTAACACTTTAAAAAGCGGATATTC	1260
Db	1201	TGTGATTTAAAAAGCATTTGAAATGAATTAACACTTTAAAAAGCGGATATTC	1260
QY	1261	AAATCTTCTTACAAATCTTGCACTTTTGATATAGATAAGCATTAATGGTGATGAA	1320
Db	1261	AAATCTTCTTACAAATCTTGCACTTTTGATATAGATAAGCATTAATGGTGATGAA	1320
QY	1321	TGTGGCGATTTATTAACCTCAAAATTTATCCATATGAAAGTTTGTATCTTTGATGTT	1380
Db	1321	TGTGGCGATTTATTAACCTCAAAATTTATCCATATGAAAGTTTGTATCTTTGATGTT	1380
QY	1381	GAGCACAATTCGGCTGGGCTTTCATTAATTAATAGAGACGAATGAATTTATATCA	1440
Db	1381	GAGCACAATTCGGCTGGGCTTTCATTAATTAATAGAGACGAATGAATTTATATCA	1440
QY	1441	TCAGCTAAACTGAATAATGATTTTAAAAAATATCTCGTTCTTATTCATTCGATCAATGG	1500
Db	1441	TCAGCTAAACTGAATAATGATTTTAAAAAATATCTCGTTCTTATTCATTCGATCAATGG	1500
QY	1501	GAAAGATTTATTCGATGAGATGAGTACATCAATCAAGAAATGAATTAAGGATTTGAT	1560
Db	1501	GAAAGATTTATTCGATGAGATGAGTACATCAATCAAGAAATGAATTAAGGATTTGAT	1560
QY	1561	AATTACACATATGAGAAATTAACCATATGATAATAAATAAAGAAATATATGATTTTCG	1620
Db	1561	AATTACACATATGAGAAATTAACCATATGATAATAAATAAAGAAATATATGATTTTCG	1620
QY	1621	AAATTAACAAAAGATATATAAAGAAAGTCACTTTAAAAATATCTGCAATGATGATCTTT	1680
Db	1621	AAATTAACAAAAGATATATAAAGAAAGTCACTTTAAAAATATCTGCAATGATGATCTTT	1680
QY	1681	TAAAGATGATGAAAAATTTATATGATGAAGGATTTGCAACTTAAGATATTTGAAAAAT	1740
Db	1681	TAAAGATGATGAAAAATTTATATGATGAAGGATTTGCAACTTAAGATATTTGAAAAAT	1740
QY	1741	CGTATGAAAAATCCTATTGATGATCCAGATGATGGCATTAAGAAATCAATTTGCGACGAA	1800
Db	1741	CGTATGAAAAATCCTATTGATGATCCAGATGATGGCATTAAGAAATCAATTTGCGACGAA	1800
QY	1801	GAAATGCGTATGCTTTCCGCTTAAAAAAAACAGCAAAAAAGAGTCT 1847 .	
Db	1801	GAAATGCGTATGCTTTCCGCTTAAAAAAAACAGCAAAAAAGAGTCT 1847	



```
RESULT 2
ABL33013/c
ID ABL33013 standard; DNA; 7597 BP.
XX
AC ABL33013;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 986.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosstatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001MO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SO Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 other;
Query Match 7.0%; Score 130; DB 24; Length 7597;
Best Local Similarity 47.7%; Pred. No. 1.2e-12;
Matches 447; Conservative 0; Mismatches 485; Indels 6; Gaps 2;
OY 912 ATACCAAGCAAAACAAATAAACAAGAAAAATTCGACGCAAAAATATTTT 971
DB 1062 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1003
OY 972 GGAATTTTAAAGGATGCTGTCACGCACTTTGCCATTTTAAACCGACTATC 1031
DB 1002 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 943
OY 1032 TTATTAAGTAAATAGATATCGTATGATTAAGTATGTTAAACAGAGTAAACA 1091
DB 942 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 883
OY 1092 ATAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1151
DB 882 TAAACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 823
```

```
OY 1152 TATATCTATTAGCTTAATATACTAAAAAGAGTAAATATGATTTGTTATTTAA 1211
DB 822 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 763
OY 1212 AAAAGCATTAGAAATGAACATTAATTAATTAATTAATTAATTAATTAATTA 1271
DB 762 AATACGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 703
OY 1272 ACAATCTTGCAATTCCTTGATTAATTAATTAATTAATTAATTAATTAATTA 1331
DB 702 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 643
OY 1332 TTTAACTCAATTTTCCATTCATGTAAGTTTGTCTTGTGAGTGGACCAATTT 1391
DB 642 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 584
OY 1392 CGGCTGGCTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1451
DB 583 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 524
OY 1452 TGAAATGATTTTAAAAAATACCTGTTTATTCATTGATCAATTTGGAGAAATTTA 1511
DB 523 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 464
OY 1512 TTACGATGAGATGATACATACATCAAGAAATGATTAAGATTAGATTAATTA 1571
DB 463 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 404
OY 1572 TGGAGAAATTAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1631
DB 403 TACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 344
OY 1632 GATATTAAGAAAGTAGCTCTTAAAAATATCGATTGATCGATTTTAAAAAGTAG 1691
DB 343 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 284
OY 1692 GAAAAATTTATTCATGATGAGATTTGCAACTTAAGATTAATTTGA-----AATATCGTATG 1746
DB 283 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 224
OY 1747 AAAATCTATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1806
DB 223 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 164
OY 1807 CCTATCTTTCCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1844
DB 163 CATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 126
RESULT 3
AAS46591/c
ID AAS46591 standard; DNA; 6286 BP.
XX
AC AAS46591;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #313.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytosstatic;
KW cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001MO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
```

	PR	06-APR-2000; 2000DE-1019058.
	PR	07-APR-2000; 2000DE-1019173.
	PR	30-JUN-2000; 2000DE-1032529.
	PR	01-SEP-2000; 2000DE-1043826.
XX	PA	(EPIG-) EPIDEMIOLOGY AG.
XX	PI	Olek A., Piepenbrock C., Berlin K;
XX	DR	WPI: 2001-602752/68.
PT	PT	Fragments of chemically modified genes associated with tumour suppressor
PT	PT	genes and oncogenes, useful in designing primers and probes for
PT	PT	analysing diseases associated with cytosine methylation state e.g.
XX	XX	cancer -
PS	Claim 1;	SEQ ID NO 313; 27pp; English.
CC	XX	The invention relates to a nucleic acid comprising a sequence of 18
CC	CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC	CC	bisulphite, of genes associated with tumour suppression and
CC	CC	oncogenes having a sequence taken from 536 (actually 533 since
CC	CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC	CC	(SS) and sequences complementary to (SS). The nucleic acid may be a
CC	CC	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC	CC	form part of a set of probes for detecting the cytosine methylation state
CC	CC	and/or single nucleotide polymorphisms and also to be used in an
CC	CC	array for analysing diseases associated with CpG dinucleotides e.g.
CC	CC	cancers and tumours. The probes can also be used in a method for
CC	CC	ascertaining genetic and/or epigenetic parameters for the diagnosis
CC	CC	and/or therapy of existing diseases or the predisposition to specific
CC	CC	diseases, by analysing cytosine methylations. The parameters may be
CC	CC	compared to another set of genetic and/or epigenetic parameters, the
CC	CC	differences serving as basis for diagnosis and/or prognosis events which
CC	CC	are disadvantageous to patients. The present sequence is one of the
CC	CC	533 genomic sequences derived from tumour suppressor genes and
CC	CC	oncogenes.
CC	CC	Note: The sequence data for this patent did not form part
CC	CC	of the printed specification, but was obtained in electronic
CC	CC	format directly from WIPO at
CC	CC	ftp.wipo.int/pub/published_pct_sequences.
XX	XX	
SQ	Sequence	6286 BP; 1508 A; 124 C; 1538 G; 3116 T; 0 other:
	Query Match	6.7%; Score 123.2; DB 22; Length 6286;
	Best Local Similarity	48.9%; Pred. No. 1.6e+11;
	Matches	362; Conservative 0; Mismatches 373; Indels 5; Gaps 14;
QY	1012 ATATTTAAACCTCAGTCATCTTTTAAGTGATATAGATATCCGTAGATTATAAAGPAT	1072
Dd	4540 ATAAATATAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAA	4488
QY	1072 GTTAAAAACGATAAAAACAATAACTTATATATTAAATCGAATTATATTGGACATGA	1133
Dd	4480 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	4422
QY	1132 TTATTATATATATAAGCATATATNCATATTTGCTAAATTATTAACAAAAGAAGCTAAA	1192
Dd	4420 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	4366
QY	1192 TATATGATTCGTATTATTTTTAAAAAGCATTAGAAAAAGAAATGAGAACATTATAAAAAGAC	1251
Dd	4360 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	4301
QY	1252 GGTCATATCAAACTCTTCTTCAATFACTTCGATTACCTTGATATATGATAAAGCATTAAT	1311
Dd	4300 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	4241
QY	1312 GGTATGATGATGGCGATATATATAACGCATATATTCATCGATGAAGAAGTTTGATCTT	1372
Dd	4240 AAAATTAATAAT-----AAATTAATTAATTAATAATAATAATAATAATAATAATAA	4186
XY	1372 CTGTAGTCTTGAGCAAAATTTGGCGCTTTCATTAATAATAATAATACAGACAGAAATGAA	1431

[illegible]





QY 1463 TTAATAAATACGTCGTTCTTATTCATTCGATCAATTGGAGAATTTATTTACGATGAGA 1522  
 Db 4107 TAAATTTAAAAAACATTCATCATATACGAAACACGAAAAATCAAAATCAATCTACC 4048  
 QY 1523 TGGATGCAATACATCAAGAAATGATTAAAGATTACATATTCACATATGGAATTA 1582  
 Db 4047 TATTCATCAATAAACCTCTAAAAACGAACTAATTAATTAATTAATTAATTAATTA 3988  
 QY 1583 CCATATGATTAATTAATTAATTAAGATATATTTGATTTGCAATTAACAAGATTTAAAGA 1642  
 Db 3987 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3928  
 QY 1643 AACTGACCTCTTAAAAATATCTGATTCGATTCGATGTTTAAAGTAGATCAAAATTTAT 1702  
 Db 3927 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3868  
 QY 1703 TGAATGAAGGATTTGCAACTAAAGATTTGAAAAATTCGTATGAAAT 1751  
 Db 3867 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3819

## RESULT 7

ABL34358/C  
 ID ABL34358 standard; DNA; 12237 BP.

AC ABL34358;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2331.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; anti-HIV; anticonvulsant; ophthalmologic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;  
 KW antihemmatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

XX 03-JAN-2002.

PD 02-JUL-2001; 2001MO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation

PS Claim 1; SEQ ID NO 2331; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 12237 BP; 3050 A; 207 C; 3024 G; 5956 T; 0 other;

Query Match 6.0%; Score 111.6; DB 24; Length 12237;  
 Best Local Similarity 48.5%; Pred. No. 1.2e-09;  
 Matches 364; Conservative 0; Mismatches 384; Indels 2; Gaps 2;

QY 1014 ATTTAAACCTGACATCTTTATAGTTATAGATATATCCGTTAGATTTAAAGTATGT 1073  
 Db 6912 AATCAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6853  
 QY 1074 TAAACGAGTAAACAAATCTTATATTTATTTGAAATTAATTTGACAGATTT 1133  
 Db 6852 AATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6793  
 QY 1134 ATTTATATTTAAGATATATCTTATAGCTTAA-ATATACATAAAAAAGAGTAAT 1192  
 Db 6792 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6733  
 QY 1193 ATATGATTTGTATTTAAAAAAGCATTTAGAAAATGAATAGACATTTAAAAAGACG 1252  
 Db 6732 AATTAATAAAAAAATAAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTA 6674  
 QY 1253 GTGATATCAATCTTCTTACATCTGATTCATCTTGATATATGATTAAGCATTTAAAG 1312  
 Db 6673 TAAATTAATAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6614  
 QY 1313 GTGATGAATGTGCGATATTTATTAACATCAATTTATCATTCATGATGAAGTTTGATCTTC 1372  
 Db 6613 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6554  
 QY 1373 TTGATGTGAGCAATTTGCGGCTTTCATTAATTAATTAATTAATTAATTAATTAATTA 1432  
 Db 6553 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6494  
 QY 1433 ATTTATCATGAGTAAACGAAATGATTTTAAAAAATACGCTTCTTATTCATTCGA 1492  
 Db 6493 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6434  
 QY 1493 TCATTTGGAGAAATTTAATTTACATGATGATGATCAATTCATCAAGAAATGATTAAG 1552  
 Db 6433 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6374  
 QY 1553 GATTAGATATTCACATATGGAATAATCACTATGATTAATTAATTAATTAATTAATTA 1612  
 Db 6373 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6314  
 QY 1613 TGAATTCGAATTAACAAAGATTAATTAAGAAAGTCAAGCTCTTAAATATCTGCATTTGAT 1672  
 Db 6313 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6254  
 QY 1673 CGATGTTTTTAAAGTGAATGAATAATTTATTTGATGAAGAGATTTGCAACTAAAGATTT 1732  
 Db 6253 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6194  
 QY 1733 GAAAAATATCGTATGAATAATCCTATTTGATGA 1762  
 Db 6193 AATTAACAAATTAATTAATCTCCATTAATCTAA 6164

RESULT 8  
 AAK73165  
 ID AAK73165 standard; DNA; 20420 BP.

XX AAK73165;

AC AAK73165;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytosine; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0244675.  
PR 08-NOV-2000; 2000US-0244676.  
PR 08-NOV-2000; 2000US-0244677.  
PR 08-NOV-2000; 2000US-0244678.  
PR 08-NOV-2000; 2000US-024523.  
PR 08-NOV-2000; 2000US-024524.  
PR 08-NOV-2000; 2000US-024525.  
PR 08-NOV-2000; 2000US-024526.  
PR 08-NOV-2000; 2000US-024527.  
PR 08-NOV-2000; 2000US-024528.  
PR 08-NOV-2000; 2000US-024532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA - ) HUMAN GENOME SCI INC.  
XX  
PA Rosen CA, Barash SC, Ruben SM;  
XX  
PI



CC skin aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis. The polypeptides can also be used as a food additive or  
 CC preservative to increase or decrease storage capabilities. The present  
 CC sequence represents a ds DNA fragment of the gene for a novel human  
 CC secreted protein of the invention.

XX Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 other;

Query Match 5.3%; Score 98.2; DB 24; Length 20420;

Best Local Similarity 48.9%; Pred. No. 1.9e-07;

Matches 346; Conservative 0; Mismatches 358; Indels 3; Gaps 3;

QY 1032 TTTATAGTAAATGATATATCCGTTAGATTATTAAGTGTAAAAACGAGTAAAAACA 1091  
 DB 4094 TATATAGTAAATGATATATCCGTTAGATTATTAAGTGTAAAAACGAGTAAAAACA 4153  
 QY 1092 ATACTATATATTTATTCGATTTATTTGACAGATATTTAATATATTAACGA 1151  
 DB 4154 TATATATTAATATATATGAAATATATATATATATATATATATATATATATATAA 4213  
 QY 1152 TATATCTATTAGCT-TAAATATTAATAAAAAAGAGTAATATATGATGTTGATTATTA 1210  
 DB 4214 TAA 4273  
 QY 1211 AAAAAAGATTAGAAAATGAAATAGACATTATTAATAAAAAAGCGTATCAATCTTCT 1270  
 DB 4274 ATATATATATATATATATATATGAAATATATATATATATATATATATATATATAT 4333  
 QY 1271 TACATATCTTCAT-TACTTGTATATAGATTAACGATTAAATGGTATGATGCGCAT 1329  
 DB 4334 ATAAATATATATATGAAATATATATATATATATATATATATATATATATATATAT 4393  
 QY 1330 ATATATTAACCTCAATTTATCCATGATGAAAGTTTGTCTTCTGATGTTGACACAAAT 1389  
 DB 4394 AT 4453  
 QY 1390 TTCGGCTGGCTTCAATAAATATATACAGACGAAATGAAATTTATATCACTTAA 1449  
 DB 4454 ATATATGAAATATATATATATATATATATATATGAAA-ATATATATATATATATATA 4512  
 QY 1450 ACGTAAATGATTTTAAAAAATACGTCCTTTATTCGATTCGATCAATGGGAAGAAATTT 1509  
 DB 4513 TATATAAATG 4572  
 QY 1510 AATTACGATGAGTACATATACATCAAGAAATGATTAAGGATTAAGATTAATACACA 1569  
 DB 4573 AAAATA 4632  
 QY 1570 TATGAGAAATATACCATATGATATATATATATATATATATATATATATATATATAT 1629  
 DB 4633 TATATTAATA 4692  
 QY 1630 AAGATATTAAGAAAGACAGCTCTTAAATATATCGATTCGATGATGTTTAAAGAGAG 1689  
 DB 4693 AATA 4752  
 QY 1690 ATGAAAAATTTATGATGAGAGGATTTGCAACTAAAGATATGAAA 1736  
 DB 4753 TATATTAATA 4799

RESULT 10

AAS46421/c ID AAS46421 standard; DNA; 13321 BP.

XX AAS46421;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #143.

KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W0200168912-A2.

PD 20-SEP-2001.

PE 15-MAR-2001; 2001WO-EP02955.

PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPICENOMICS AG.

P1 Olek A, Piepenbrock C, Berlin K;

DR WPI; 2001-602752/68.

XX Claim 1; SEQ ID No 143; 27bp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18

CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and

CC oncogenes having a sequence taken from 536 (actually 533 since

CC numbers 408, 458 and 500 are missing from the sequence listing) sequences

CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a

CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may

CC form part of a set of probes for detecting the cytosine methylation state

CC and/or single nucleotide polymorphisms and also to be used in an

CC array for analysing diseases associated with CpG dinucleotides e.g.

CC cancers and tumours. The probes can also be used in a method for

CC ascertaining genetic and/or epigenetic parameters for the diagnosis

CC and/or therapy of existing diseases or the predisposition to specific

CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the

CC differences serving as basis for diagnosis and/or prognosis events which

CC are disadvantageous to patients. The present sequence is one of the

CC 533 genomic sequences derived from tumour suppressor genes and

CC oncogenes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 13321 BP; 2515 A; 580 C; 4121 G; 6105 T; 0 other;

Query Match 5.3%; Score 97.8; DB 22; Length 13321;

Best Local Similarity 47.6%; Pred. No. 2.2e-07;

Matches 288; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

QY 1035 ATAGTAATATGATATATCCGTTAGATTATTAATATGTTAAACGAGTAAACACATA 1094

DB 2152 ATACGTTAAATATACATTCAGCTTATTAATACAACTTCCTCTTAATAATAATAA 2093

QY 1095 ACTATATATTTTATTCGATTTATTTTGCACGTGATTTATTTATATATTAAGACATAT 1154

DB 2092 ATAAATATCATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2093

QY 1155 ATCTATTAGCTTAAATTAATACTAAAAAAGAGTAAATATATGATGTTGTTATTTAAAA 1214

DB 2032 ACAATATTAACGAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1973

QY 1215 AGCAATTGAAATATGAAATAGACACTTTATAAAAAAGACGCGATATCAAACTTCTTACA 1274















GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 02:19:36 ; Search time 2425.86 Seconds  
(without alignments)  
11082.453 Million cell updates/sec

Title: US-08-876-132-1

Perfect score: 1660  
Sequence: 1 AGACTACACAAAGCAAAATT.....AAGACCTCGAGCAGATCT 1660

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estimu:\*  
5: em\_estlov:\*  
6: em\_estpl1:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl1:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_estl3:\*  
13: gb\_estl4:\*  
14: gb\_estl5:\*  
15: em\_estfun:\*  
16: em\_estlom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
c 1	118	7.1	879 17	CNS01JRG
c 2	115.6	7.0	1101 17	CNS0021J
c 3	110.8	6.7	619 9	AL514935
c 4	110	6.6	1036 17	CNS031MJ
c 5	109.6	6.6	1025 17	CNS014J2
c 6	109	6.6	1036 17	CNS00599

c 7	109	6.6	1135 17	CNS033GQ	AL226115 Tetradon
c 8	106	6.4	924 13	BM415947	BM415947 OP21029 M
c 9	104.2	6.3	1223 17	B12981	B12981 F2AD11-Sp6
c 10	103.2	6.2	1885 10	BE420745	BE420745 HRM002.B0
c 11	103	6.2	964 13	BM416130	BM416130 OP21217 M
c 12	102.8	6.2	910 13	BM415636	BM415636 OP20714 M
c 13	102.2	6.2	907 17	CNS021J4	AL176953 Tetradon
c 14	101.2	6.1	974 17	CNS001TT	AL1765432 Tetradon
c 15	100.6	6.1	934 13	BM415088	BM415088 OP20159 M
c 16	100.4	6.0	948 13	BM415494	BM415494 OP20570 M
c 17	100.4	6.0	998 17	CNS00YXQ	AL096968 Drosophila
c 18	99.6	6.0	966 13	BM415686	BM415686 OP20766 M
c 19	99.4	6.0	982 13	BM415348	BM415348 OP20422 M
c 20	99.2	6.0	990 17	CNS0060I	AL065652 Drosophila
c 21	98.2	5.9	937 17	CNS0066L	AL062959 Drosophila
c 22	98.2	5.9	1101 17	CNS00EVL	AL069706 Drosophila
c 23	97.2	5.9	1139 17	AO897537	AO897537 HS_3153-A
c 24	96.6	5.8	660 17	BH183498	BH183498 023_L_07-
c 25	96.6	5.8	660 17	CNS070NJ	AL620449 T3 end of
c 26	96.6	5.8	804 17	AG077527	AG077527 Pan trogl
c 27	96.4	5.8	976 17	AG136173	AG136173 Pan trogl
c 28	95.8	5.8	1055 14	B0876453	B0876453 AGENCOURT
c 29	94.6	5.7	1101 17	CNS0039R	AL063932 Drosophila
c 30	94.2	5.7	838 17	AG061324	AG061324 Pan trogl
c 31	94.2	5.7	896 13	BM415641	BM415641 OP20719 M
c 32	93.6	5.6	811 17	CNS03YOO	AL266649 Tetradon
c 33	93.6	5.6	960 17	AG031602	AG031602 Pan trogl
c 34	93.6	5.6	1101 17	CNS00KRB	AL077453 Drosophila
c 35	93.4	5.6	834 17	B12387	B12387 F21E20-Sp6
c 36	93.4	5.6	864 17	CNS04NBY	AL298519 Tetradon
c 37	93.2	5.6	1059 17	CNS0022B	AL097133 Drosophila
c 38	93	5.6	1300 13	BM468018	BM468018 AGENCOURT
c 39	92.6	5.6	1185 12	BE273407	BE273407 GAL_Eb001
c 40	92.4	5.6	915 12	BG520365	BG520365 ps20C09.Y
c 41	92.2	5.6	959 17	CNS00655	AL062806 Drosophila
c 42	91.8	5.5	1079 17	CNS0351X	AL228786 Tetradon
c 43	91.6	5.5	815 17	AG044049	AG044049 Pan trogl
c 44	91.6	5.5	1309 13	BM463041	BM463041 AGENCOURT
c 45	91.4	5.5	1021 17	AG032747	AG032747 Pan trogl

## ALIGNMENTS

RESULT 1  
CNS01JRG/LOCUS 879 bp DNA linear GSS 12-JUN-2001  
DEFINITION Anopheles gambiae GSS T7 end of clone 14D07 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.  
ACCESSION AL147405  
VERSION AL147405.1 GI:7005551  
KEYWORDS African malaria mosquito.  
SOURCE ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
REFERENCE 1 (bases 1 to 879)  
Genoscope.  
Genoscope.  
Direct Submission  
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
2 (bases 1 to 879)  
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.  
Direct Submission  
Submitted (16-FEB-2000) BBMT, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France  
This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.

# FEATURES

## Source

Location/Qualifiers  
1. 879  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="14D07"  
/clone\_lib="Notredame1"  
/note="end : 17"

BASE COUNT 66 a 38 c 1 g 670 t 104 others

## ORIGIN

Query Match 7.1%; Score 118; DB 17; Length 879;

Best Local Similarity 42.8%; Pred. No. 1.8e-10;

Matches 356; Conservative 48; Mismatches 420; Indels 7; Gaps 2;

```

QY 16 AAAAAAATAGATTAATTTTCGAGATTAATTAAGCCGACTTAACAATGAGTG 75
DB 849 AAAAAAAAAAAAAAAAAAAAAAAAAAGTAGTAATATTTTGAATGTAATTAATA 790
QY 76 AAGAAGAAAGAAAAAATAATACATTTTGGCTAGTAAGAGAAAGAAAAATPA 135
DB 789 AAAAAAAAAAGAAAAAGTAAATTAATAATATGTAAGAAKAAWAAAAATATA 730
QY 136 AAGAAGACTCGCTTAACGTCAAGAAACGAAATATATAAAGAAAGAGACTGTGAT 195
DB 729 GTAAATAAATATGACAMADKAKKAAAAAGTAAKAGWAAAAAATAAAAAA 670
QY 196 TTTTAATGGAATCGTGAGAAAGAAATTTTATTTTCATTTTCAGAGGATTAATTG 255
DB 669 AAGKATATATAAAGAAAAAAGAAAAAAGRTTATKMKAMKAGAAAAAAGKGA 610
QY 256 TTGTAGTTGATGAAAAATCTAGATTAATAATGACATCAAAATGTGTTGAATTGCA 315
DB 609 AAAAAATGGKDAADAAAYTTGAAAAAAGAGAGWAGGKRGARAAAGAAAGKAK 550
QY 316 TTATTGAATACGTATATCAATTAATGGGGGTTGTCTATTTATTTTGGAGATTTG 375
DB 549 AAAAAAAAAAATGTTAATAATGAAAAAAGKATRRAAAAAAMKAGDAAAKRGMAG 490
QY 376 AAAAACTGAGTGAAGAAATAGTTTGGAGAGCAAAAAACCTGCCGTTTTCATCA 435
DB 489 AAAAAAGATWKAKAAAAA-----AAAAATTAATAAATKAMTWBAAAAAAGARAA 435
QY 436 ATGACTTTGGAATAATTCATTTGTGAGCGGTAGCAAACTTGAATTTTTCATTTGA 495
DB 434 GKATFATKTDGRRAAAAAAADAANANAAAAAATAAAAAAATAAAAAA 375
QY 496 AATTGAAAAAATAAGCAAAAAAGAACTCAATGCAAAAAATATATATATAAAGAGA 555
DB 374 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 315
QY 556 GATCGATATGATTTTAAAGCAGAAACGACATTTGAATGAAAAAAGATTTGGAAA 615
DB 314 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 255
QY 616 AATATCTATGCGAGTAATTAAGCAAAAAAATTTGGAACTCAACCCGGTGTGTT 675
DB 254 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 195
QY 676 TAGAATGAGCATGAAAGAAATGATGAATAATATCAACCTGATTTAATGAAGAAAG 735
DB 194 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 137
QY 736 CAGGCAATATAGCAAAATTTATCAAAATTAAGTTGAGCATAGTAATACAGATGATG 795
DB 136 AAAAAAATTAATAAATAATATTAATAAATAAANNNNNNNNNNNAANCAABETTTTANN 77
QY 796 TAACGGAGCTATAGAGTGTGGAACCTGCAAAATAGTTTGTAAACAAGAA 846
DB 76 AANNANNNNTATNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 26

```

RESULT 2

## CNS0021J

## LOCUS

CNS0021J 1101 bp DNA linear GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Oseogawa and  
Aaron Mammosser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR05N11"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

## BASE COUNT

631 a 7 c 28 g 289 t 146 others

## ORIGIN

```

QY 7 ACACAGCAAAATGAAAAATAGATAAATTTTCGAGGTATTAAGCCGACTTAAGC 66
DB 302 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 361
QY 67 AATGAGTGAAGAGAAAAAATAAATACATATTTTGAATTTTGAATTAAGAGAG 126
DB 362 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 421
QY 127 AAAAAAATGAAGAGACCTCGCTTAACAGTGAAGAAACGAAATATAAAGAGAA 186
DB 422 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
QY 187 GACTGTGATTTTAAATGGAATCGTGAGGAAAGAAATTTTAATTTTCATTTGAGGG 246
DB 482 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 541
QY 247 ATTAATTTGTGTAAGTGAAGAAATCTGATTAATAAATGCGATCAAAAATGTGTG 306
DB 542 TTTTATTTTATTTTAAATTTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 601
QY 307 AATTTGACATTTATGAAATAGTATATCAATTAATGAGGGGTGTGCTATTTATTTG 366
DB 602 WTTTAAATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 658
QY 367 CGAAGATTTGAATAATCTGAGTGAAGAAATAGTTTGGAGAGCAAAAAACCTTGGCCT 426

```





REFERENCE 3 (bases 1 to 1036)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000)  
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
 source  
 1..1036  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0B036CG06SP1-end ; PUC-ORI"

BASE COUNT 1005 a 6 c 7 g 17 t 1 others  
 ORIGIN

Query Match 6.6%; Score 110; DB 17; Length 1036;  
 Best Local Similarity 44.8%; Pred. No. 3.8e-09;  
 Matches 422; Conservative 0; Mismatches 520; Indels 0; Gaps 0;

```

OY 11 AAGCAAAATGAAAAATGATTAATTTCCAGGTATTAAGCCGACTTAACAAAT 70
DB 38 AAGCAAAAAAATGAAAAATGATTAATTTCCAGGTATTAAGCCGACTTAACAAAT 97
OY 71 GAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130
DB 98 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 157
OY 131 AATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 190
DB 158 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 217
OY 191 GCGATTTTAATGGAATCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 250
DB 218 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 277
OY 251 ATTTGTTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGT 310
DB 278 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 337
OY 311 TGACATTATGAAATACGTATATCAATTAATGGGGCTTCTATTATTGCGGAA 370
DB 338 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 397
OY 371 GATTGAAATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 430
DB 398 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
OY 431 TTCAATGACTTTGGAAAAATTCATGTGACGCGTAGCGAACTTTGAAATTTTACA 490
DB 458 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
OY 491 TTGGAATTTGAAATTAAGCAAAAGAACTCAAAATGGAATTAATTAATTAATA 550
DB 518 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 577
OY 551 AAGGAGATCGATGATGATTTTAAAGCAGAACTGACATGATGAAATTAATTAAT 610
DB 578 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
OY 611 GGAATAAATCTATGCTGAGATGATTAATTAAGCAAAATTTGGAACTCAACCGGTGT 670
DB 638 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
OY 671 TGTTTAGAAATGACATGAAGAGAAATGATGAATATCAACCTCGATGTTAATGAAGA 730
DB 698 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
OY 731 AAGCAGCTCAATATAGGAATTTTCAAAATTAATAGTGACATAGTAATCGAGATGA 790
DB 757 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
  
```

DB 758 AA 817

OY 791 TCTAGTACGAGACTTATAGTGTGAGACTGCAATAGTTTGAATAAACAAGATGC 850

DB 818 AA 877

OY 851 CTTTGGTTTGTATTTGAGAGAAATTAAGCAACTGAGAAAGAACTGATTAATCAAG 910

DB 878 AA 937

OY 911 AAGAGTAAAGATTCGATACATGGAACCAAACTAAGA 952

DB 938 AA 979

RESULT 5  
 CNS014J2/c  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN1111 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL104216  
 VERSION AL104216.1 GI:5615827  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1025)

REFERENCE  
 AUTHORS Direct Submission  
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [sequef@genoscope.cns.fr](mailto:sequef@genoscope.cns.fr))  
 JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

COMMENT  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeLoBac11.

FEATURES  
 source  
 1..1025  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="BACN1111"  
 /clone\_lib="DrosBAC"  
 /plasmid="pBeLoBac11"  
 /note="end : 77"

BASE COUNT 148 a 88 c 6 g 522 t 261 others  
 ORIGIN

Query Match 6.6%; Score 109.6; DB 17; Length 1025;  
 Best Local Similarity 35.0%; Pred. No. 4.4e-09;  
 Matches 283; Conservative 98; Mismatches 427; Indels 0; Gaps 0;

```

OY 69 ATGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128
DB 890 RTTGGDRARADAAAAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 831
OY 129 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
DB 830 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771
OY 189 CTGATTTTAAATGAGAAATCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 248
DB 770 KRTTTRTGTGTAATAAATATGTRGTATATATATATATATATATATATATATAT 711
OY 249 TAAATTTGTTGATGATGAGAAATCTAGATTAATTAATGAGATCAAAATGCTGGA 308
DB 710 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 651
  
```



LOCUS CENS0360 1135 bp DNA linear GSS 15-MAY-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ORI end of clone  
 208924 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 ACCESSION AL226115.1 GI:7885026  
 VERSION AL226115.1  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 1 (bases 1 to 1135)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fizames,C., Mincker,P., Brotlier,P., Quetier,F.,  
 Saurin,W. and Weissenbach,J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence  
 Unpublished  
 2 (bases 1 to 1135)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 Unpublished  
 3 (bases 1 to 1135)  
 Genome scope.  
 Direct Submission  
 Submitted (12-APR-2000)  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.  
 FEATURES  
 Source  
 1. 1135  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="208P24"  
 /clone\_1lb="G"  
 /note="Genoscope sequence ID : C0AG208DH12SP1-end :  
 PUC-ORI"  
 BASE COUNT 863 a 65 c 43 g 124 t 40 others  
 ORIGIN  
 Query Match 6.6%; Score 109; DB 17; Length 1135;  
 Best Local Similarity 44.6%; Pred. No. 5.4e-09;  
 Matches 358; Conservative 14; Mismatches 431; Indels 0; Gaps 0;  
 Oy 7 ACACAAAGCAATTCGAAAAATAGATAAATTTTCGACGATTAAGCCGACTTAAAC 66  
 Db 216 AA 275  
 Oy 67 AAATGAGGAG 126  
 Db 276 AA 335  
 Oy 127 AAAAAAAAAAG 186  
 Db 336 AA 395  
 Oy 187 GACTGTGATTTTAATGGAATTCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
 Db 396 AA 455  
 Oy 247 ATTATTTTGTGTAGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
 Db 456 AA 515  
 Oy 307 AATTGACATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 366  
 Db 516 AA 575

Oy 367 CGAGATTGAATCTGAGTGAAGAAATAGTTTCGAGAGACAAAAAACCCTTGCCGT 426  
 Db 576 AA 635  
 Oy 427 TTTTTCGAATGACTTTGGAAAAAATTCATGTGTGCGGTAGCGAACTTGAAATTTT 486  
 Db 636 AA 695  
 Oy 487 TACATTGGAATTTTAAAAAATTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
 Db 696 AA 755  
 Oy 547 AAAAAAGAGATTCGATTCGATTTTAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
 Db 756 AA 815  
 Oy 607 ATTTCGAAAAAATCTATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
 Db 816 AWACAAAAAC 875  
 Oy 667 GTGTGTTTGAATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
 Db 876 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 935  
 Oy 727 AAGAAACAGCAGGTCATATAGAGAAATTTTCAGAAATTAAGTTGAGCATAGTAATCAG 786  
 Db 936 AACCAA 995  
 Oy 787 ATGATCTAGTACGGGACTATTA 809  
 Db 996 ACACATTAATAAAAAAAAAAAAAAAAAATATA 1018

## RESULT 8

BM415947/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM415947 924 bp mRNA linear EST 28-JAN-2002  
 OP21029 Mixed Stage EST's from Globodera pallida, the potato cyst  
 nematode Globodera pallida CDNA, mRNA sequence.  
 BM415947  
 BM415947.1 GI:18382746  
 EST.  
 Globodera pallida.  
 Globodera pallida  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
 Tylenchoidea; Heterodermidae; Heteroderinae; Globodera.  
 1 (bases 1 to 924)  
 Heer,J., Sosinski,B., Pokrzywa,R.M., Warry,A. and Opperman,C.  
 Mixed Stage EST's from Globodera pallida, the potato cyst nematode  
 Unpublished (2001)  
 Contact: Opperman, C  
 Center for the Biology of Nematode Parasitism  
 NC State University; IACR-Rothamsted  
 Campus Box 7616; Raleigh, NC 27695, USA  
 Tel: 919.515.6699  
 Fax: 919.515.9500  
 Email: warthog@unity.ncsu.edu  
 GT11-8PCN\_R\_F10\_GT11-8\_R\_079\_ab1.  
 FEATURES  
 Source  
 1. 924  
 /organism="Globodera pallida"  
 /db\_xref="taxon:36090"  
 /clone\_1lb="Mixed Stage EST's from Globodera pallida, the  
 potato cyst nematode"  
 /note="Vector: lambda GT11: This is a collaborative effort  
 between IACR-Rothamsted and North Carolina State  
 University. The library was constructed from mixed stage  
 G. pallida in lambda GT11 by Paul Burroughs,  
 IACR-Rothamsted."

BASE COUNT 55 a 49 c 10 g 796 t 14 others  
 ORIGIN  
 Query Match 6.4%; Score 106; DB 13; Length 924;

Best Local Similarity 45.2%; Pred. No. 1.8e-08;  
Matches 361; Conservative 0; Mismatches 438; Indels 0; Gaps 0;

```

OY 1 AGATCTACACAGCAATTTGAAAAATAGATTAATTTTGGCAGGTATTAAGCCGACT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 807 AAAAAAAGGAGAGAAAAATAAAAAAGAAAAAAGAAAAAAGAAAAAATATAA 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 61 TAAACAAATGATGAGAGAAAAAAGAAAAATTAATATTTTGTAGTTAGTAAAG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 747 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAATTAATTAAGAAAA 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 121 AGAAAAAATATTAAGAGAGCTCGGCTTAACAGTCGAAAAACAGAAATTAATAAA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 AAAAAAAGGAGAGAAAAAAGAAAAAATTAAGAAAAAAGAAAAAATTAAGAAAA 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 181 GAAAGAGCTGATTTTATGCAATCGAGAGAAAAAAGAAATTTTATTTTCAATTT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 627 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 241 CGAGGATTAATTTGTTAGTTGATGAAAAATCTAGTAAAAATGAGATCAAAAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 567 AAAAAAAGGAGAGAAAAAAGAAAAAATTAAGAAAAAAGAAAAAATTAAGAAAA 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 301 GTGTTGAATTTGACATTTATGAAATACGATATATCAATATGAGGTTTGTCTATT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 507 TAAAAAATTAATTAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAATTA 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 361 ATTTTGCAGATTGAAAAATCTGAGTGAAGAAATAGTTTGGCAGAGCAAAACCT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 447 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 421 TGCGCTTTTTCATTAATGCTTTGAAAAAATTCATTTGAGCGGTAGAGAACTTGA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 481 ATTTTTCATTTGAAATTTTGAAGAAATTAAGCAAGAACTCAATGAGAAAAAT 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 327 NNTTGGGGGGTAAATTAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 541 ATTATTAAGAAAGAGATCGATATGATTTTAAAGCAAGAACTGACATTGAAGA 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 601 AAAAAATTTGAAAAATCTATGCTGAGTGAATTAAGCAAAATTTGGGAGCTC 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 661 AACCGGTGTTTGTAGAAATGAGATGAAGAAATGATGAAATATCAACTCGATG 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 147 ACTAAGAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 721 TTAATGAAGAAAGAGAGCTCAATATGAAATTTTCAAAATTAAGTTGAGCATGA 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 87 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 28
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 781 AATCAGATGATCTACTAAC 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 27 AACAGNTGAGCTCGATC 9
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9  
B12981/c 1223 bp DNA linear GSS 14-MAY-1997  
LOCUS T24D11-Sp6 Arabidopsis thaliana genomic clone T24D11, DNA  
DEFINITION sequence.  
ACCESSION B12981  
VERSION B12981.1 GI:2094103  
KEYWORDS GSS.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1223)  
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.  
TITLE BAC End Sequences at ATCC  
JOURNAL Unpublished (1997)  
COMMENT Contact: Ecker J.  
Arabidopsis thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
19104  
Tel: 215-898-9384  
Fax: 215-898-8780  
Email: jecker@genome.bio.upenn.edu  
Seq primer: Sp6  
Class: BAC ends  
High quality sequence start: 199  
High quality sequence stop: 271.  
Location/Qualifiers  
1..1223  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T24D11"  
/clone\_lib="TAMU"  
/sex="hermaphrodite"  
/note="Vector: BeLoBACII; Site\_1: HindIII; Site\_2: HindIII  
; Produced by Rod Wing"

BASE COUNT 70 a 50 c 53 g 678 t 372 others  
ORIGIN

Query Match 6.3%; Score 104.2; DB 17; Length 1223;  
Best Local Similarity 30.5%; Pred. No. 3.4e-08;  
Matches 295; Conservative 0; Mismatches 672; Indels 0; Gaps 0;

```

OY 7 ACACAAGCAATTTGAAAAATAGATTAATTTTGGCAGGTATTAAGCCGACTTAAC 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1220 AANNAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 67 AATGAGTGAAGAGAAAAAAGAAAAATTAATATCATTTTGTAGTTAGTAAAGAAAG 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1160 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 127 AAAAAATTAAGAGAGCTCGGCTTAACAGTCGAAAAACAGAAATTAATTAAGAGA 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1100 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 187 GACTGTGATTTTATGAAATCGTGAAGAAAGAAATTTTATTTTCAATTTTGAAGG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1040 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 247 ATTAATTTTGTAGTATGAAGAAATCTAGATTAAGAAATGCAAGTCAAAATGTCTG 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 980 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 307 AATTGACATTTTGAATAGCTAGTATATCAATATGAGGTTTGTCTATTATTATTG 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 920 AANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 367 CGAAGATTGAAGATCTGAGTGAAGAAATAGTTTGCAGAGCAAAAAACCTTGGCGT 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 860 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 427 TTTTTCATATGACTTTGAAAAAATTCATTTGTGAGCGGTACGGAACCTTGAATTTT 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 800 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 487 TACATTTGAATTTTGAAGAAATTAAGCAAGAACTCAATGAGAAAAATTAATTATA 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 740 NANNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 547 AAAAAAGAGATCGATATGATTTTAAAGAGCAAGAACTGACATTTGAATGAAGAAAG 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 680 AAAAAAAGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 607 ATTGGAAAAATCTATCTGAGCTGAAATTAAGCAAAAAATGGGAACCTCAACCG 666  
 Db 620 AA 561  
 QY 667 GTGCTGTTTGAATGACGATGAAAGAAATGATCAAAATATCTCACTCGATGTTAATG 726  
 Db 560 AA 501  
 QY 727 AAGAAACAGCAGCTCATATATAGAAATTTATTCAAAAATTAAGTGAATAGTAATCAG 786  
 Db 500 AA 441  
 QY 787 ATGATCTAGTAACGGGACTATTAGAGTGTGAACTGGAATGATTGATTAACAGAA 846  
 Db 440 NNN 381  
 QY 847 GTGCTTCTGTTTGTATTTGTGTGAGAAATTCACCACTGAGAAAGAGCTGATATG 906  
 Db 380 NNN 321  
 QY 907 CAGAGAGTAAGATTCATACATGAAAGCAAAAGCTAAAGAGCTTTGAAATGA 966  
 Db 320 NNN 261  
 QY 967 GTTTTGT 973  
 Db 260 ATTTTAT 254

## -RESULT 10

BE420745 1885 bp mRNA linear EST 24-JUL-2000  
 LOCUS HMW002.B02 ITEC HMM Barley Leaf Library Hordeum vulgare cDNA clone  
 DEFINITION HMW002.B02, mRNA sequence.  
 BE420745  
 BE420745.1 GI:9418588  
 EST.  
 SOURCE Hordeum vulgare.  
 ORGANISM Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 1885)

REFERENCE  
 AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,  
 S., Dubcovsky,D., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,  
 Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,  
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,  
 Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,  
 Sorrells,M., Warburton,M. and Wenzel,G.  
 International Triticeae EST Cooperative (ITEC): Production of  
 Expressed Sequence Tags for Species of the Triticeae  
 Unpublished (2000)

JOURNAL  
 COMMENT Botantes Institut der LMU  
 Contact: Hermann RG  
 Menzinger Str. 67, D-80638 Munchen GERMANY  
 Fax: 49 30 171683  
 Email: hermann@botanik.biologie.uni-muenchen.de  
 International Triticeae EST Cooperative (ITEC)  
 http://wheat.pw.usda.gov/genome.

FEATURES  
 source  
 1..1885  
 /organism="Hordeum vulgare"  
 /cultivar="Barke"  
 /db\_xref="taxon:4513"  
 /clone="HMW002.B02"  
 /clone\_1ib="ITEC HMM Barley Leaf Library"  
 /tissue\_type="Leaf"  
 /dev\_stage="14 day old"  
 /note="Vector: pbluescriptsk(-): 850 bp average insert  
 size."

BASE COUNT 1138 a 219 c 212 g 176 t 140 others  
 ORIGIN

Query Match 6.2%; Score 103.2; DB 10; Length 1885;  
 Best Local Similarity 42.4%; Pred. No. 4,4e-08;  
 Matches 384; Conservative 0; Mismatches 521; Indels 0; Gaps 0;

QY 48 ATTAAAGCCGACTTAAACCAATGATGAGAGAGAAAGAAATTAATCATATTTT 107  
 Db 967 AANNNAANNAANNN 1026  
 QY 108 GAGTGTAGTAAAG 167  
 Db 1027 AA 1086  
 QY 168 GAAATTAATTAAG 227  
 Db 1087 AA 1146  
 QY 228 TAATTTTCAATTTTCAGAGGATTAATTTGTGTGATGATAAATCTAGTTAAAT 287  
 Db 1147 AA 1206  
 QY 288 GCAGATCAAAATGCTGTAATTTGACATTTGAAATACGATATATATATGCGG 347  
 Db 1207 AA 1266  
 QY 348 GTTGTCTATTTTATTTTTCGAGAGATGAAATCTGACTGAAAGAAATAGTTGCGAGA 407  
 Db 1267 AA 1326  
 QY 408 GCAG 467  
 Db 1327 AA 1386  
 QY 468 GCGAAGCTTGAAATTTTTCATTTGAAATTTGAAATTAAGCAAGAACTCAA 527  
 Db 1387 AA 1446  
 QY 528 TGGAAAAATTTATTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587  
 Db 1447 AA 1506  
 QY 588 ACATTTGATGAG 647  
 Db 1507 AA 1566  
 QY 648 AATTTGGAAGCTCAACCGGTTGTTTACAAATGACAGTAAGAAATGAGAAAT 707  
 Db 1567 AA 1626  
 QY 708 ATCAAGCTCGATGTTATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 767  
 Db 1627 AA 1686  
 QY 768 GTTGACATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 827  
 Db 1687 AA 1746  
 QY 828 AGTTTGTAAAG 887  
 Db 1747 AA 1806  
 QY 888 AGAAGAGAGCTGATATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 947  
 Db 1807 AA 1866  
 QY 948 AAGCA 952  
 Db 1867 AAAAA 1871

RESULT 11  
 BM416130/c 964 bp mRNA linear EST 28-JAN-2002  
 LOCUS BM416130

DEFINITION	OP21217 Mixed stage EST's from Globodera pallida, the potato cyst nematode					
ACCESSION	BM416130					
VERSION	BM416130.1 GI:18382929					
KEYWORDS	EST.					
SOURCE	Globodera pallida.					
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.					
REFERENCE	1 (bases 1 to 964)					
AUTHORS	Heer,D., Sosinski,B., Pokrzywa,R.M., Warrly,A. and Opperman,C.					
TITLE	Mixed Stage EST's from Globodera pallida, the potato cyst nematode unpublished (2001)					
JOURNAL	Contact: Opperman, C Center for the Biology of Nematode Parasitism NC State University; IACR-Rothamsted Campus Box 7616; Raleigh, NC 27695, USA Tel: 919.515.6699 Fax: 919.515.9500 Email: warthog@unity.ncsu.edu GRIIPCNL_F06_1-30R_047.abl.					
COMMENT	Location/Qualifiers					
FEATURES	1..964					
SOURCE	/organism="Globodera pallida" /db_xref="taxon:36090" /clone_lib="Mixed Stage EST's from Globodera pallida, the potato cyst nematode" /note="Vector: lambda GT11; This is a collaborative effort between IACR-Rothamsted and North Carolina State University. The library was constructed from mixed stage G. pallida in lambda GT11 by Paul Burroughs, IACR-Rothamsted."					
BASE COUNT	42 a	36 c	18 g	852 t	16 others	
ORIGIN						
Query Match	6.2%	Score 103;	DB 13;	Length 964;		
Best Local Similarity	44.5%;	Pred. No.5.8e-08;				
Matches	424;	Conservative	0;	Mismatches 524;	Indels 5;	Gaps 1;
Dd	123	AAAGCAAAAAAATGAAGAAGACCTCGGCTTAACACTCGAAAACCAATAATATAAAGA	182			
Dd	963	AAATATAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAAAAAA	904			
Oy	183	AAGAGACTGTGATTTTAAATGAAATCGTGAGAAACAATAATTTTCAATTTTCG	242			
Dd	903	AAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	844			
Oy	243	AGGATTAATTTGTTGTAAGTTGATGAAAAATCTAGATTAATAAATGCAGATCAAAATGT	302			
Dd	843	AGAAAAAATAAAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAA	784			
Oy	303	GTTGAATTTGACATTATGAATFACGTATCATCAATATGGGGGTTTGCTTTTTAT	362			
Dd	783	AAAAAATATAAAAAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAA	724			
Oy	363	TTTGGGAGATGAAATCTGAGTGAAGAATAATAGTTTGGAGAGCAAAAAACCCTTG	422			
Dd	723	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	664			
Oy	423	CCGTTTTTTCAATGACTTTGGAAAAAATTCATTGTGAGCGGTGACCAACTTTGAAT	482			
Dd	663	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	609			
Oy	483	TTTTTACATGGAATTTGAATAAATAAAGCAAAAGAACTCAATATGGAATAATATAT	542			
Dd	608	AAAAAAGAGAAATTA	549			
Oy	543	TATTA	602			
Dd	548	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	489			
Oy	603	AAAGATTTGGAAAAATCTATGCTGAGAGTGAATTAAGCAAAAAAATTTGGAACTCA	662			

[illegible]







FEATURES  
sourceLocation/Qualifiers  
1. .934

/organism="Globodera pallida"  
/db\_xref="taxon:36090"  
/clone\_idb="Mixed Stage EST's from Globodera pallida, the  
potato cyst nematode"  
/note="Vector: Lambda GT11; This is a collaborative effort  
between IACR-Rothamsted and North Carolina State  
University. The library was constructed from mixed stage  
G. pallida in Lambda GT11 by Paul Burroughs,  
IACR-Rothamsted."

BASE COUNT 45 a 80 c 17 g 777 t 15 others  
ORIGIN

Query Match 6.1%; Score 100.6; DB 13; Length 934;  
Best Local Similarity 45.6%; Pred. No. 1.5e-07;  
Matches 360; Conservative 0; Mismatches 429; Indels 1; Gaps 1;

```
QY 9 ACAAGCCAAATTGAAAAATAGATAAATTTTCGACGATATTAAAGCCGACTTAAACAA 68
DB 792 ATAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 733
QY 69 ATGACGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128
DB 732 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 673
QY 139 AAAAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
DB 672 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 613
QY 189 CTGTCATTTTAAATGGAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 248
DB 612 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 553
QY 249 TAAATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308
DB 552 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
QY 309 TTTGACATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 368
DB 492 TAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 433
QY 369 AAGATTTGAAATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
DB 432 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 374
QY 429 TTTTCAATGACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
DB 373 GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
QY 489 CATTGGAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
DB 313 GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
QY 549 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
DB 253 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
QY 609 TTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
DB 193 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 134
QY 669 GTTGTTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
DB 133 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 74
QY 729 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 788
DB 73 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14
QY 789 GATCTAGTAA 798
DB 13 TATCAGCTAA 4
```

Search completed: July 17, 2003, 06:38:51  
Job time : 244.36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 02:19:35 ; Search time 79.9943 Seconds

(Without alignments)  
6363.996 Million cell updates/sec

Title: US-08-876-132-1

Perfect score: 1660  
Sequence: 1 AGATCTACACAGCAAAATT.....AAGGACTCCGACGACATCT 1660Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PCBUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	81.2	4.9	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	69.6	4.2	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 3	67.8	4.1	5852	1 US-07-867-106-2	Sequence 2, Appl
C 4	62.6	3.8	615	4 US-08-998-416-186	Sequence 166, App
C 5	60.2	3.6	1956	4 US-08-559-896B-1	Sequence 1, Appl
C 6	59	3.6	6243	2 US-09-056-075-1	Sequence 1, Appl
C 7	57.6	3.5	837	4 US-08-998-416-288	Sequence 288, App
C 8	57	3.4	240	1 US-08-628-417-6	Sequence 6, Appl
C 9	56.6	3.4	2570	2 US-09-056-075-2	Sequence 2, Appl
C 10	55.8	3.4	636	4 US-08-998-416-1137	Sequence 1137, Ap
C 11	55.4	3.3	356	2 US-08-520-678A-22	Sequence 22, Appl
C 12	55.4	3.3	356	4 US-08-897-126-22	Sequence 10, Appl
C 13	55.4	3.3	9646	3 US-08-811-566-1	Sequence 1, Appl
C 14	55.4	3.3	9646	4 US-09-034-756-1	Sequence 1, Appl
C 15	54	3.3	658	4 US-08-998-416-595	Sequence 595, App
C 16	54	3.3	2447	2 US-09-014-969-14	Sequence 14, Appl
C 17	53.8	3.2	8920	2 US-08-446-855A-1	Sequence 1, Appl
C 18	53.8	3.2	8920	4 US-09-150-741-1	Sequence 1, Appl
C 19	53	3.2	4766	5 PCT-US93-07261-10	Sequence 10, Appl
C 20	52.8	3.2	3275	4 US-09-370-838-151	Sequence 151, App
C 21	52.2	3.1	6152	4 US-08-973-462-1	Sequence 1, Appl
C 22	51.6	3.1	12980	3 US-08-811-566-5	Sequence 5, Appl
C 23	51.6	3.1	12980	4 US-08-034-756-5	Sequence 5, Appl
C 24	51	3.1	1798	4 US-09-797-906-1	Sequence 1, Appl
C 25	50.8	3.1	2030	4 US-09-512-342-1	Sequence 1, Appl
C 26	50.6	3.0	72604	4 US-09-268-992-7	Sequence 7, Appl
C 27	50.6	3.0	72604	4 US-09-657-474-7	Sequence 7, Appl

28	50.2	3.0	2674	4 US-09-817-180-1	Sequence 1, Appl
29	49.4	3.0	2223	1 US-08-257-073-4	Sequence 4, Appl
C 30	49.2	3.0	3718	4 US-09-424-283-6	Sequence 6, Appl
C 31	48.8	2.9	6265	4 US-09-129-112-3	Sequence 3, Appl
C 32	48.8	2.9	13737	4 US-09-538-414-10	Sequence 10, Appl
C 33	48.6	2.9	20674	4 US-09-641-638-651	Sequence 1, Appl
C 34	48.4	2.9	51952	3 US-08-947-823-1	Sequence 1, Appl
C 35	48	2.9	1493	1 US-08-340-820-24	Sequence 24, Appl
C 36	48	2.9	1193	1 US-08-593-535-24	Sequence 33, Appl
C 37	47.8	2.9	1117	4 US-09-247-373B-33	Sequence 2, Appl
C 38	47.8	2.9	5852	1 US-07-867-106-2	Sequence 13, Appl
C 39	47.8	2.9	19124	2 US-08-487-826B-13	Sequence 9, Appl
C 40	47.6	2.9	3000	1 US-08-764-100-9	Sequence 1, Appl
C 41	47.6	2.9	168575	4 US-09-426-290-1	Sequence 1, Appl
C 42	47.4	2.9	394	2 US-08-623-906A-7	Sequence 7, Appl
C 43	47.4	2.9	1066	1 US-08-157-101A-4	Sequence 4, Appl
C 44	47.4	2.9	3188	1 US-08-017-664-4	Sequence 27, Appl
C 45	47.2	2.8	1447	4 US-09-443-041A-27	

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-F1s  
US-08-232-463-14  
Query Match 4.9%: Score 81.2: DB 1: Length 7218:

Best Local Similarity 2.28; Pred No. 1.6e-08;  
Matches 8; Conservative 241; Mismatches 119; Indels 0; Gaps 0;

OY	480	AATTTTTCACATTGGAAATTTGAAAAAATAAGCCAAAAGAACTCAATGTGA AAAAATAT	539
Dd	1444	AATTTGGACRRR	1387
OY	540	TATTATATAAAAAGAGATCGGATATGATTTTAAGCAGAAAACTGACATTGAATGAA	599
Dd	1384	RR	1322
OY	600	AAAAAAGATTGGAAAAAATCTAGCTGAGACTGATTAAGCAAAAAAATTTGGAACT	659
Dd	1324	RR	1265
OY	660	CACOCGGTGTTTGTGAAGAATGCAGTAAGAAAGATGAAGAAAATATCAOCTGCAT	719
Dd	1264	RR	1205
OY	720	GTTATGAGAAGACAGAGTCAATATATGAGAAATATCAAAAAAAGTTAGACATAGT	779
Dd	1204	RR	1145
OY	780	AAATCAGATGATCTAGTAACGGACTATTAGAGTGTGGAAGCTCGAATAAGTTTGCATAAA	839
Dd	1144	RR	1085
OY	840	ACAAGAG 847	
Dd	1084	RRRRRRRR 1077	

RESULT 2  
US-08-487-826B-13  
Sequence 13, Application US/08487826B

GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelssen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid

```

; STRANDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

```

Query Match	4.28	Score 69.6	DB 2	Length 19124
Best Local Similarity	47.08	Pred. No. 5.8e-06		
Matches 287	Conservative 0	Mismatches 314	Indels 9	Gaps 2

QY	2	AAAAATAGATAAAAATTTTCCAGAGATAGAGCCGCTTAAACCAATGAGTGGAGAGAG	82
Db	15360	ATAATATGGAAGTCTGGTATAGTGTATATATATATATATATATATATATATATATATTA	15419
QY	83	AAAGAAAAAATATTAATCATATTGTTGAGTACTAGTAAGAAAGAAAGAAAAATATAACAGA	142
Db	15420	AAAAAGAAATATTAACCAATTTATTTAAATGAAAAAAGAAAAATGAAATATATAAAAA	15479
QY	143	CCTGGCTTACAGTCGAAAAACCGAATATATAAAAAGAAAGACTGTGATTTTAAT	202
Db	15480	AAATTTTATTAATTAATAAAAAAAGAGAAAAATTTTAAATAATAT	15539
QY	203	GGAATTCGTAGGAAAAAGAAATTTTAAT---TTTCATTTTCGAGCATTAATTGTG	258
Db	15540	AAAAATATATTAATATTAATTTTGATGATTAATAATGAAAAAGTTTTCAAAAA	15599
QY	259	TAACTGATCAAAAATCTAGATTAATAAATGCAGATCAAAAATGTGTGAATTTGACATTA	318
Db	15600	AAATTTAAAAAATTTTATATAAAAAAGATTATAAAAAATTAACCAAAAA	15659
QY	319	TTGAAATACGTACTATATCATTAATGAGGGCTTGTCTATTTTATTTGCGAAGATTGAA	378
Db	15660	AGAAAAAACAATTAATAAAAAAATAATATATCATTAATAACAAAAAAGAAA	15719
QY	379	-----ATCTAGAGAAAGAAATAGTTTGGGAGACAAAAAACCTTCCGCTTTTTC	433
Db	15720	AAATATATTAATTAATAATATATATATATATATATTAATAATTAATAAATTTAAAAA	15779
QY	434	AAATGACTTTGAAAAAATTCATTGAGCGGAGGAGAACTTTGAAATTTTTCATTG	493
Db	15780	AAAAAATATATCATTAATAATAAAAAATTTTATTAATTAATAAAAAATATTAATTA	15839
QY	494	GAAATTTGAAAAATAGCGCAAAAGAACTCAATGAAAAAATATATATTAATAAAG	553
Db	15840	AAATTTTAATTAATTAATAAAAAATATTAATTAATAAAAAATTTAATTAATAAAAAAT	15899
QY	554	GAGATCGATATGATTTTAAAGACGAATACTGACATTTGAATGAAAAAAGATTTGGA	613
Db	15900	TAAAAAATTTATGAAATTAATAAAAAAATTAATAATTAATAAAAAAATTAATA	15959
QY	614	AAAAATCTAT	623
Db	15960	TAAAAATTAAT	15969

RESULT 3  
 US-07-867-106-2/C  
 : Sequence 2, Application US/07867106  
 : Patent No. 5389526  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Slade, Martin B  
 : APPLICANT: Chang, Andy C M  
 : APPLICANT: Williams, Keith L  
 : TITLE OF INVENTION: Improved Plasmid Vectors for Cellular  
 : TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium  
 :  
 : NUMBER OF SEQUENCES: 19  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris  
 : STREET: One Liberty Place 46th Floor  
 : CITY: Philadelphia  
 :  
 : STATE: PA  
 :  
 : COUNTRY: USA

Query Match	4.1%	Score 67.8	DB 1	Length 5852
Best Local Similarity	46.3%	Pred. No. 1e-05		
Matches 258	Conservative 0	Mismatches 297	Indels 2	Gaps 1
Qy	49	TTAAAGCCGACTTAAACCAAAATGAGTGAAGAAAGAAAACAAAATAATTAATCATATTTTG	108	
Db	2341	TTAAATTAATAATTAATTAATGAATTTTCATTAATTAATTAATTAATTAATGAAGTGTG	228	
Qy	109	AGTTAGTAAAGAGAGAGAAAAATTAAGAAGACCTCGGCTTAACAGTCGAAAAACAG	168	
Db	2281	GGAAAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA	222	
Qy	169	AAATATATAAAGAAAGAGACTGTCATTTTAAATGAAATCGTGAGGAAAAAGAAATTTT	228	
Db	2221	AAAAAATAAATAATTAATCAAAAAATTAATAATTAATTAATTAATTAATTAATTAATTA	2162	
Qy	229	AATTTTCATTTTCGAGGATTAATTTGTATAGTTGATGAAAAATCAGATTAATAAATG	288	
Db	2161	ATATTTGATATATTAATCCATATATTTATATAATAGGTATAATTAGTAAAGATTAATAATTT	2107	
Qy	289	CAGATCAAAAAATGCTGTAATTTGACATTATGAAATCGATATATCAATTAATGAGGG	348	
Db	2101	TAAATTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2042	
Qy	349	TTTGTCTATTTTATTTTCCGAAGATGAAAAATCTGATGGAAGAAATAAGTTGCGAG	408	
Db	2041	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAAGTAAAAAATAAATAAATAAATAA	1982	
Qy	409	CAAAAAAACCCCTGCGCTTTTTTTCAAATGACTTGAAAAAATTT--CATGTGACGGT	466	
Db	1981	AAATRGAAAAAAGTTGGTTAAACTACATCATAGTTTATATAGTTTTCATATTTAAAT	1922	
Qy	467	AGCGAAACCTTGAAATTTTATACATGTGAAATTTGAAAAATATAGCAAAAGAACTCA	526	
Db	1921	AACCTTTATTTTAATGATTTTAAATATGAGAGTCATATAAAAAATAAATAAATTTTAA	1862	
Qy	527	ATGAAAAAATATTTTATTAATAAATAAAGAGATCGGATATGATTTTAAAGCAAGAACT	586	

```

RESULT 4
US-08-998-416-186/C
Sequence 186 Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippson, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebschuhg, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264rtlis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998.416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: pr/5-30306/A/CGCI976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8669
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAGI074RP
US-08-998-416-186
Query Match 3.8%; Score 62.6; DB 4; Length 615;
Best Local Similarity 46.0%; Pred.No.7.6e-05;
Matches 247; Conservative 0; Mismatches 289; Indels 1; Gaps 1;
Oy 61 TAAATCAATGAGCGAAGAAGAAAAAATAATACATATTGGTACTAGTAAGA 120
Db 548 TTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 489
Oy 121 AGAAGAAAATATTTAAAGAAAGAACCTCGGCTTAACAGTCGAAGAAACCAAGAAATTAATAAAA 180
Db 488 AGAAATTTAAATTTAAATTAATTTTAATTAATTAATTTCTTATTAAGAAAGTAAATTAATTAATA 429
Oy 181 GAAGAGACGTGTGATTTTTTAATVGAAGAAATCGTGAGGAAAGAAAATTTTAATTTTCATTTT 240

```

```

Db      428 TCAACATATATTTATTAATAAGATATTTATTAATAATAATAATTTTCAATATTTTAA 369
Qy      241 CGAGGATTAATTTGCTGTAGTGTGAATAATCTAGATTAATAATGAGATCAAAAT 300
Db      368 TAAATTAATCTTTATTAATAATAATATTTATTTTAAACAATTAATAATATAT 309
Qy      301 GTGTGAATTTGACAT-TATTGAAATACGTACTATATCAATATATGGGGTTGTCTATTT 359
Db      308 ATTAATATTTGATTAATCTTTTAAATATTTTAAAGAAATATATATCTAATATATAT 249
Qy      360 TTTTTCGCAAGATGAATAATCTGAGTGAAGAAATAGTTCCGAGACCAAAAAACC 419
Db      248 TTTAATACTAATTTTAAATTTGAACATGACTAAATAGTATTCATTTAAATATTTT 189
Qy      420 TTGCGCTTTTTCCTCAATGACTTTGAAAAAATTCATTTGACCGGTAGCAAACTTGA 479
Db      188 TTATATATTAATAATATTAATAATGATGATTAATTAATTAATTAATATATATAT 129
Qy      480 AATTTTTCATTTGCAAAATTTGAAAAATTAAGCAAAAGAACTCAATGAAAAATAT 539
Db      128 AAGTATTTAATCAATATTTAATTTAATAATATGATTAATATAGTTTAAATATTT 69
Qy      540 TTTTATTAATAAGAGATCGATATGATTTTAAAGCAGAAACTGCATTTGAT 596
Db      68 AATACTTAATATTTTATTAATAATAAGCTTATATTAATCTTTAATAATTAATATAT 12

```

## RESULT 5

US-08-559-896B-1

Sequence 1, Application US/08559896B

Patent No. 6310046

GENERAL INFORMATION:

APPLICANT: Patrick E. Dufly

APPLICANT: Christian F. Ockenhouse

TITLE OF INVENTION: SEQUESTIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: John Moran

STREET: USA MRMC - MCMR-JA

CITY: PORT DETRICK, FREDERICK

STATE: MARYLAND

COUNTRY: USA

ZIP: 21702-5012

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 7.5

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/559, 896B

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Moran, John

REGISTRATION NUMBER: 26, 313

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 619-2065

TELEFAX: (301) 619-7714

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1956 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

US-08-559-896B-1

Query Match 3 6%; Score 60.2; DB 4; Length 1956;  
 Best Local Similarity 41.9%; Pred. No. 0.00032;

	Matches	Conservative	0;	Mismatches	598;	Indels	6;	Gaps	1;
Qy	25	AAATAGATAAATTTTCCGAGTATTTAAAGCCGACTTAAACAAATGAGTGAAGAGAA	84						
Db	884	ATAGAGAAAAATTTACCAAAATTTGAAAAAGAGAAAAATTAATAATGATTAAGATCAAA	943						
Qy	85	AGAAAAAATTAATACATTTTGTGACTTGTAAAGAGAGAAAAAATTAAGAGAACCC	144						
Db	944	TGATTAATAATATATGAGAGAAATTAACAAAATGATGTGATGAAATTCACATGTAA	1003						
Qy	145	TCGCGTACAGTCGAAAAACCAAAATTAATAAGAAAGAGACGTGATTTTATG	204						
Db	1004	GGAGACCAATTTACAGATATACAAAAAGAAAAATTAACAAATTTAGACTTGAAGAAA	1063						
Qy	205	AAATCGTGAGAAAAAATTTTAAATTTTCATTTTTCGAGGATTAATTTGTGTAAGTT	264						
Db	1064	TCGATGACTTTTAAAGCAATTTGATGATGATGATGATGATGATGATGATGATGATG	1123						
Qy	265	GATGAAAAATCTAGATTAATAATGCAATCAAAATGCTGTGAATTTGACATTTGAAA	324						
Db	1124	CCATGGAATTTTAAAGCAAAATGAAAAAGATTAATTTTACATGAAATTTAAAAATG	1183						
Qy	325	TACGTAGTATATCAATATGAGGGGTTTGTCTATTTTATTTTTCGAAATTTGAAATCTGA	384						
Db	1184	AATCTTACCAAAAAATTAAGAGAAAAATGTAATGATTTTATTAATACGACAAATATG	1243						
Qy	385	GTGAAAAAATATGTTTTCGAGAGCAAAAAACCCCTGCGGTTTTCGAAATGACTTTG	444						
Db	1244	ATAGTATATATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA	1303						
Qy	445	GAAAAAATTCATTTGAGCGGTAGCAAACTTTGAAATTTTATTTTACATTTGAAATTTGAAA	504						
Db	1304	CAAAATTAATTAATTAATTAATGAAATGTAAGAGTTGAATTTGATGCAAAATTTAGCAAAAG	1363						
Qy	505	AAATAGGCAAAAGCAAACTCAATGCAAAAAATTAATTAATTAATAAGAGATCGATA	564						
Db	1364	ATTAAGAGCCCAAGATAGAGATTTATGACTATTTTAAACAAGAAATTTAAAAAGACA	1423						
Qy	565	TGATTTTAAAGCAGAAACTGACATTTGATGAAAAAAGATTTGAAAAAATCTATG	624						
Db	1424	AAAAATGTAATGTTTCCCAATATAGTGAATTTTAAATTTCAAAAGATGAGAAAGATTAACA	1483						
Qy	625	CTGAGAGTAAATTAAGCAAAAAAATTTGGAACTCAACCGGTGTTTGAAGAATGA	684						
Db	1484	CACCAATTCACATTAAGAAAGAAATCAATGATGATGTTCTCAGAAAAATTTACGATTA	1543						
Qy	685	CGATGAAGAAATGATGAAAAATATCAACCTGATGTTTAATGAGAAACAGCAGTCAAT	744						
Db	1544	TTCAAGAGATATATTAATAATAATAAGCCCAAGAGATTAACACTGAATGTTAGATTAATA	1603						
Qy	745	ATAGCAATTTATCAAAAAATTAAGTTGACATATGTAATCAATGATCTAGTAGCGGAC	804						
Db	1604	ATTAAGAAATTAACAAATATTTGATTAATAAAAAATTTGATGATTAATAAAAAATTTGGAGATA	1663						
Qy	805	TATTAGAGTGTGAACCTGAAATGTTTGAATAAACAAGAGTGCCTTTCTTTTGA	864						
Db	1664	TAAAAAGTGTGAGATTAATAAAGTGTGATGATTAATAACATGTTGATGATTAATAA	1723						
Qy	865	TTTGAGAGAAATTCAGCACT-----GAGAAAGAGCTGATTAATGCAAGAGATA	918						
Db	1724	ATGTTGATGATTAATAAATTTGATGATTAATAAATTTGATGATTAATAAATTTGATG	1783						
Qy	919	AAGATTTGATCAATTAAGAGCAAAACTTAAGAGGCTTTGAATTTGATTTGTTTGG	978						
Db	1784	GAGATATTAACAATCTGAGATTAACAATAATGCTGAGATTAACAATTTGAGGAGATA	1843						
Qy	979	ATAAGATTTTGTGAGTGAATAATGCAATTAATGATGATATTTCTCAACAAAAAAG	1038						
Db	1844	TAAACAATCTGATATATATCAACGTTGAACATATAGACGAGCGGAAAAAACCA	1903						
Qy	1039	ACTCTGCAAGTAAAGAA 1057							
Db	1904	ATCTTGATTAATCCAAAAA 1922							

RESULT 6  
US-09-056-075-1  
Sequence 1, Application US/09056075  
Patent No. 5953368  
GENERAL INFORMATION:  
APPLICANT: Johnson, Eric A.  
APPLICANT: Bradshaw, Marite  
APPLICANT: Rood, Julian  
TITLE OF INVENTION: Expression System for Clostridium  
TITLE OF INVENTION: Species  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,075  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95238  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 3770..4013  
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from  
OTHER INFORMATION: plasmid RP4"  
US-09-056-075-1

Query Match 3.6%; Score 59; DB 2; Length 6243;  
Best Local Similarity 51.7%; Pred. No. 0.00076;  
Matches 134; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
DB 9 ACNAGGCAATGAAATATAGATTAATTTTCGCGAGTATTAAGCCGCTTAACAA 68  
1192 AAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1251  
QY 69 ATGATGAG 128  
DB 1252 ATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1311  
QY 129 AAAATTAAG 188  
DB 1312 AAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1371  
QY CTGTGATTTTATGAG 248  
DB 1372 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1431  
QY 249 TAATTTGTTGTAAGTTGAT 267  
DB 1432 TTTTATTAATTAATTAAT 1450

RESULT 7  
US-08-998-416-288/C  
Sequence 288, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NO. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: P6G1241RP  
ORGANISM: P6G1241RP  
US-08-998-416-288

Query Match 3.5%; Score 57.6; DB 4; Length 837;  
Best Local Similarity 44.8%; Pred. No. 0.00093;  
Matches 260; Conservative 0; Mismatches 319; Indels 1; Gaps 1;  
DB 18 ATTGAATAATTAATTAATTAATTTTCGCGAGTATTAAGCCGAGTAAACAAATGAGTGA 77  
591 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 532  
QY 78 GAAG 137  
DB 531 AATGAG 472  
QY 138 GAAG 197  
DB 471 TAATTTTATTAATTAATTTCTTAATTAAGAGTTAATTAATTAATTAATTAATTA 412  
QY 198 TTAATGAG 257  
DB 411 AATATGAG 352

OY 258 GTAAGTGTGATAAATCTAGATATAAATGACATCAAAATGCTTGAATTGCAT- 316  
DB 351 ATAAATAAATAATATATATTTATTAATAACATTAATAAATAATATATATTTATGATATC 292  
OY 317 TATGAAATAGTATATATCAATATATGGGGTTTGTCTATTTTATTTGGCAAGATGCA 376  
DB 291 TATTTAAATAATTTATTAAGAAATAATAATATCTAATAATTTAATACTAATTTAA 232  
OY 377 AAATCGAGTGAAGAAATAGTTGCGAGAGCAAAACCCCTGCGTTTTTTTCAAA 436  
DB 231 AATTGACATGACTAATATAGTATTCATTTAAATATATTTATATATATATAAATA 172  
OY 437 TGACTTGGAAAAATTCATTTGTGAGCGGTAGCGAAACTTTGAAATTTTACATTTGAA 496  
DB 171 TTAATAATGATGAATTAAGTAAATATATTAATTAATATATATATATATATCAAA 112  
OY 497 ATTGCAAAAAATAGCGCAAAACCTCAATGCAAAATATTTTATTAATAAAGAGG 556  
DB 111 TAAATTAATTTTATTAATATGATATAATAGTTTAAATAATTTAATCTTAAATTTATA 52  
OY 557 ATCGATATGATTTTAAAGCAGAAACTGACATTTGAAT 596  
DB 51 ATAAAAAGTTTATTTAATCTTTTAAATTAATTAATATAT 12

RESULT 8  
US-08-628-417-6

; Sequence 6, Application us/08628417  
; Patent No. 5627054  
; GENERAL INFORMATION:  
; APPLICANT: GILLESPIE, DAVID  
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC  
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL  
; ADDRESS: DEFENSE COMMAND  
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)  
; CITY: ABERDEEN PROVING GROUND  
; STATE: MARYLAND  
; COUNTRY: USA  
; ZIP: 21010-5423  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628.417  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BIFFONT, ULYSSES J  
; REGISTRATION NUMBER: 39,908  
; REFERENCE/DOCKET NUMBER: DAM 398-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 410-671-1158  
; TELEFAX: 410-671-2534  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: oligodeoxynucleotide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: YES  
; US-08-628-417-6

Query Match 3.4%; Score 57; DB 1; Length 240;  
Best Local Similarity 54.5%; Pred. No. 0.00092;  
Matches 114; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

OY 16 AAATTGAAAAATAGATATAAATTTTTCGAGGTATTTAAAGCCGACTTAAACAAATGACTG 75  
DB 12 ACAATTAACCTTGTGAATTAATTTTACTAAAAAATAAATAAATAAATAAATAAATAA 71  
OY 76 AAGAAGAAAAAATAAATAATATATATTTTGTAGTTGTAAGAGAAAAAATAAATA 135  
DB 72 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 131  
OY 136 AAGAAGCCTGCGCTTACAGTCGAAAAACCAATAATATAAAGAAAGAGACTGTGAT 195  
DB 132 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 191  
OY 196 TTTTAATGAAATCGTGAGAAAGCAAA 224  
DB 192 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 220

## RESULT 9

US-09-056-075-2  
; Sequence 2, Application us/09056075  
; Patent No. 595368

; GENERAL INFORMATION:  
; APPLICANT: Johnson, Eric A.  
; APPLICANT: Bradshaw, Marile  
; APPLICANT: Rood, Julian  
; TITLE OF INVENTION: Expression System for Clostridium  
; TITLE OF INVENTION: Species  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056.075  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95238  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2570 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-056-075-2

Query Match 3.4%; Score 56.6; DB 2; Length 2570;  
Best Local Similarity 44.8%; Pred. No. 0.002;  
Matches 218; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

OY 170 AATAATAAAGAAAGAGACTGTGATTTTAAATGAATCGTGAGAAAAAATAATTTTA 229  
DB 116 AACATATACATATCAATATTTTGTAAACCTAATAATATATATCAAAATTTT 175  
OY 230 ATTTTCATTTTCGAGGATTAATTTGTTAGTTGATGAATAAATCTGATTAATAAATGC 289  
DB 176 ATTAGATGTTTACATATATGATTAATTTTGTGTAATAAATGCGCTTATTAATAATTTA 235



```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAGI692RP
; US-08-998-416-1137

Query Match      3.4%; Score 55.8; DB 4; Length 636;
Best Local Similarity 44.7%; Pred. No. 0.0021;
Matches 257; Conservative 0; Mismatches 317; Indels 1; Gaps 1.

QY 23 AAAAAATGATAAATTTTCCGACGATTTAAAGCCGACTTAAACAAATGATGAGAAAGA 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 AAAAAATGATGATTTAAATTTAAATTAATTAATTTAAATTAATTAATTAATTAATTAAT 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 83 AAGAGAAAAAATTAATACATATTTTGAGTTGTAAGAAAGAAAGAAAAATTAAGAAGA 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 GAATTTAATTTATTTAAATATTAATAATTTAATTAATTAAGAAATTTAAAGTTAAATTAAT 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 143 CCTCGGCTTAACAGTCGAAAAACAGAAATTAATATAAAGAAAGACAGCTGTGATTTTAAT 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 TTAAATTAATTTCTTATTAAGATTTAATTAATTAATTAATTAATTAATTAATTAAT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 203 GGAATTCGTGAGAAAGAAAAATTTTAATTTTCATTTTCGAGGATTAATTTGTGTAAAG 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 AGAATTTAATTAATTAATTAATTTTCAAAATTTTAATTAATTAATTAATTTAATTAAT 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      346  AAAATATATATTTTAAATAAACAATTAAATAATATATATTAATTAATCATATATCATTT 287
QY      322  AAATACGATATATCAATTAATGGGGTTTGTCTATTTTATTTTGGCAAGATTGAAATC 381
Db      286  AATATATTTTAAAGAAAATATATATATCTATATATATTTTAAATCAATTTTAAATTT 227
QY      382  TGAGTGAAGAAATATAGTTTCGAGAGCAAAAAAACCCCTGGCGTTTCTTTCAAATGACT 441
Db      226  GAACATAGACCTAAATAGTATTCATTTAAATATTTTATTAATATTAATTAATATTAAT 167
QY      442  TTGCAAAAAATTCATTGTGAGCGGTAGCGAAACTTTGAAATTTTTCATTGGAAATTTG 501
Db      166  AATGATGATTAATAGTAATATATATATATATATTAATAAGATTTAAATCAATTAAT 107
QY      502  AAAAAATAGCAAAAGAACTCAAAATGCAAAAAATTTATATTAATAAAAAAGATCGG 561
Db      106  AATTTATTTAAATTAATGATTAATATATGTTTAAATTTAATTAATCTTAATATTAATAA 47
QY      562  ATATGATTTTAAAAAGCAAAAACTGACATTGAAT 596
Db      46  AAGTTTATATTAATCTTTAATAATTAATATAT 12

RESULT 11
US-08-520-678A-22/c
; Sequence 22, Application US/08520678A
; Patent No. 5874565
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles M.
; APPLICANT: Kolykhalov, Alexander A.
; TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
; TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howell & Haferkamp, L.C.
; STREET: 7733 Forsyth Blvd., Suite 1400
; CITY: St. Louis
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,566  
FILING DATE: 03-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 1113-1-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 9646 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-811-566-1

Query Match 3.3%; Score 55.4; DB 3; Length 9646;  
Best Local Similarity 56.2%; Pred. No. 0.0049;  
Matches 104; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 10 CAAGGCAAAATTGAAAAATAGATAAATTTTCGAGGTATTAAAGCCGACTTAAACAA 69  
DB 9585 CACAGCTAGCCGCTAGCTAGGCTTAAGATGAGCCACCATTTAAAGAAAGAAAGAA 9526  
QY 70 TGAAGTGAAG 129  
DB 9525 GGAAG 9466  
QY 130 AATATAAG 189  
DB 9465 AAGAGAGCC 9406  
QY 190 TGTGA 194  
DB 9405 TAAGA 9401

RESULT 14  
US-09-034-756-1/c  
Sequence 1, Application US/09034756  
Patent No. 6392028

GENERAL INFORMATION:  
APPLICANT: RICE, CHARLES et al.  
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C VIRUS (HCV) AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOWELL & HAFERKAMP, L.C.  
STREET: 7733 FORSYTH BLVD., SUITE 1400  
CITY: ST. LOUIS  
STATE: MO  
COUNTRY: USA  
ZIP: 63105

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/034,756  
FILING DATE: 04-May-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HOLLAND, DONALD R.  
REGISTRATION NUMBER: 35,197  
REFERENCE/DOCKET NUMBER: 6029-4831  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 314-727-5188  
TELEFAX: 314-727-6092

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 9646 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-034-756-1

Query Match 3.3%; Score 55.4; DB 4; Length 9646;  
Best Local Similarity 56.2%; Pred. No. 0.0049;  
Matches 104; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 10 CAAGGCAAAATTGAAAAATAGATAAATTTTCGAGGTATTAAAGCCGACTTAAACAA 69  
DB 9585 CACAGCTAGCCGCTAGCTAGGCTTAAGATGAGCCACCATTTAAAGAAAGAAAGAA 9526  
QY 70 TGAAGTGAAG 129  
DB 9525 GGAAG 9466  
QY 130 AATATAAG 189  
DB 9465 AAGAGAGCC 9406  
QY 190 TGTGA 194  
DB 9405 TAAGA 9401

RESULT 15  
US-08-998-416-595/c  
Sequence 595, Application US/08998416  
Patent No. 6239264

GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Redischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 6239264artis Corporation  
STREET: 3054 Cornwalis Road  
CITY: Research Triangle Park  
STATE: NO. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 595:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 658 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1408RP

US-08-998-416-595



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 02:22:20 ; Search time 612.027 Seconds  
(without alignments)  
5635.000 Million cell updates/sec

Title: US-08-876-132-1

Perfect score: 1660  
Sequence: 1 AGATCTACACAGCAAAATT.....AAGGACTCGGACGAGATCT 1660

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 144851 seqs, 1038787357 residues

Total number of hits satisfying chosen parameters: 2897702

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS07\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq3.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.8	5.0	9539	US-10-239-676-52	Sequence 52, Appl
2	83.6	5.0	7823	US-10-239-676-198	Sequence 198, Appl
3	74.2	4.5	18997	US-10-172-086-17	Sequence 17, Appl
4	71	4.3	9539	US-10-239-676-51	Sequence 51, Appl
5	70	4.2	411	US-09-960-352-14521	Sequence 14521, A
6	69.4	4.2	446	US-09-960-352-3400	Sequence 3400, Ap
7	69	4.2	821	US-09-878-574-4653	Sequence 4653, Ap
8	69	4.2	960	US-10-198-846-6381	Sequence 6381, Ap
9	67.4	4.1	327	US-09-960-352-4630	Sequence 4630, Ap
10	67.4	4.1	7657	US-10-239-676-185	Sequence 185, Appl
11	66.6	4.0	428	US-09-960-352-573	Sequence 573, Appl
12	65.4	3.9	12405	US-10-239-676-35	Sequence 35, Appl
13	64.8	3.9	4985	US-10-094-240-10	Sequence 10, Appl
14	64.2	3.9	5689	US-10-239-676-90	Sequence 90, Appl
15	64.2	3.9	6098	US-10-239-676-126	Sequence 126, Appl
16	63.4	3.8	529	US-09-983-965-2109	Sequence 2109, Appl

C 17	63.4	3.8	6665	14	US-10-239-676-4	Sequence 4, Appl1
C 18	63.2	3.8	525	14	US-10-198-846-1483	Sequence 1483, Ap
C 19	62.8	3.8	6352	14	US-10-172-086-24	Sequence 24, Appl
C 20	62.8	3.8	1223197	14	US-10-027-632-179264	Sequence 179264, A
C 21	62.4	3.8	6306	14	US-10-239-676-224	Sequence 224, Appl
C 22	62.2	3.7	5314	14	US-10-155-533-1	Sequence 1, Appl1
C 23	62	3.7	15832	14	US-10-239-676-106	Sequence 106, Appl
C 24	61.4	3.7	408	11	US-09-960-352-1221	Sequence 1221, Ap
C 25	61.2	3.7	516	11	US-09-960-352-5785	Sequence 5785, Ap
C 26	61	3.7	479	11	US-09-960-352-12872	Sequence 12872, A
C 27	60.8	3.7	344	11	US-09-960-352-1036	Sequence 1036, Ap
C 28	60.8	3.7	17848	14	US-10-239-676-28	Sequence 28, Appl
C 29	60.6	3.7	424	11	US-09-960-352-11218	Sequence 11218, A
C 30	60.6	3.7	5127	14	US-10-239-676-132	Sequence 132, Appl
C 31	60.4	3.6	390	11	US-09-960-352-3640	Sequence 3640, Ap
C 32	60.4	3.6	431	11	US-09-960-352-5558	Sequence 5558, Ap
C 33	60.4	3.6	4237	10	US-09-745-763-20	Sequence 20, Appl
C 34	60.2	3.6	1267	14	US-10-001-843-45	Sequence 45, Appl
C 35	60.2	3.6	1267	15	US-10-001-843-45	Sequence 45, Appl
C 36	60.2	3.6	1956	10	US-09-351-794A-1	Sequence 1, Appl1
C 37	60.2	3.6	7657	14	US-10-239-676-185	Sequence 185, Appl
C 38	59.8	3.6	2000	11	US-09-938-842A-3403	Sequence 3403, Ap
C 39	59.8	3.6	15732	14	US-10-239-676-96	Sequence 96, Appl
C 40	59.6	3.6	451	11	US-09-960-352-10262	Sequence 10262, A
C 41	59.6	3.6	640681	11	US-09-790-988-1	Sequence 1, Appl1
C 42	59.4	3.6	6815	14	US-10-239-676-50	Sequence 50, Appl
C 43	59.4	3.6	11047	14	US-10-239-676-187	Sequence 187, Appl
C 44	59.4	3.6	12465	14	US-10-239-676-31	Sequence 31, Appl
C 45	59.2	3.6	283	11	US-09-960-352-9095	Sequence 9095, Ap

## ALIGNMENTS

RESULT 1  
US-10-239-676-52/c  
: Sequence 52, Application US/10239676  
: Publication No. US20030082609A1  
: GENERAL INFORMATION:  
: APPLICANT: OLEK, Alexander  
: APPLICANT: BERLIN, Kurt  
: TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
: FILE REFERENCE: 5013.1003  
: CURRENT APPLICATION NUMBER: US/10/239, 676  
: CURRENT FILING DATE: 2002-09-24  
: PRIOR APPLICATION NUMBER: PCT/EP01/03968  
: DE 10019058.8  
: DE 10019173.8  
: DE 10032529.7  
: DE 10043826.1  
: PRIOR FILING DATE: 2001-04-06  
: 2000-04-06  
: 2000-04-07  
: 2000-06-30  
: 2000-09-01  
: NUMBER OF SEQ ID NOS: 228  
: SEQ ID NO 52  
: TYPE: DNA  
: ORGANISM: Artificial Sequence  
: FEATURE:  
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-52

Query Match 5.0%; Score 83.8; DB 14; Length 9539;  
Best Local Similarity 43.0%; Pred. No. 3.2e-05;  
Matches 463; Conservative 0; Mismatches 612; Indels 2; Gaps 1;

QY 4 TCACACAGCAAAATGAAATAGATTAATTTTCGAGGATTAAGCGAGCTAA 63  
DB 1116 TCTTCTAAAAAAGCAAAACGAAAAAATTAATAAATAAATAA 1057

OY		64	AACAAATGAGCAACAGAAAAAATAAATATCATTTTGTGTAAGTAAAGACA	123
Dd		1056	AA	997
OY		124	AAGAATAATTAAMAGAACCTCGGCTTAACAGTCGAAAACACAGAAATTAATAAGAA	183
Dd		996	TAAAAAATTAAAAAATAAATTAATTAATAAATAAATAAATAAATAATAT	937
OY		184	AGAGCTGTGATTTTTATGGAATCGTGAGAAAAAGAAAATTTAATTTTCATTTCGA	243
Dd		936	AAACAAAAAATAAATAAACGGAAATTAATTAACACACTATATCCCAATCATTAAAAA	877
OY		244	GGAATTAATTTGTTSTAAGTTCATGAAAAATCTAGATTAATAAATAACAGATCAAAATGHC	303
Dd		876	ACTAATAATTAATAAATAATCACTTAACCAAATTTCAAAACCAACTTAACCAACAAATCA	817
OY		304	TTGAATTTGACATTATTTGAATACGTAGATATATCAATAATGGGGTTGTCTATTATT	363
Dd		816	TACCCCAATTTCTACAAAAAATAAATACTAAACCAACCTTAATACACACACTATPA	757
OY		364	TTGCAAAATTTGAAATCTGAGTGAAGAAATTAATTTGCCGAGACAAAAACCCTGC	423
Dd		756	TCCCACTACTCAAAAAAATAAACAAAAAATACTACCGGAAAAACAATAATTACAAATA	697
OY		424	CGTTTTTTCAATACCTTTGGAAAAAATTCATTTGTGACGGGTAGCGAACTTTGAAT	483
Dd		696	AACGAAATCGACACTACACTCCACACTTAACAACAAAAAACCCCTACCTCAAAAAA	637
OY		484	TTTTTCATTTGGAATTTGAAAAAATTAAGCAAAAGAACTCAATTGAAAAATTTATT	543
Dd		636	AATATATTTAATTAATAAATAAATAAATAATATATTAATAAATAAATAAATAAATA	577
OY		544	ATAAAAAAGGATCGGATATGATTTTAAAGCAAGAAATCGATTAATGAATAA	603
Dd		576	AAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATC	517
OY		604	AAGATTTGAAAAAATCATGCTGAGAGTGAATTAAGCAAAAAAATTTGGAACTCAAC	663
Dd		516	AAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	457
OY		664	CCGGTGTGTTTAGAATGACGATGAAGAAATGATGAATAATATCAACCTCGATGTTA	723
Dd		456	AAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	397
OY		724	ATGAAGAAGACGACGATCAATATAGAAATTTTCAAAATTAAGTTGAGCATAGTAAT	783
Dd		396	AAAAAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	337
OY		784	CAGATGATCTAGTAACGGGACTATTAGAGTGTGAATCTCGAAATAGTTTGATTAACA	843
Dd		336	AATAAAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACA	277
OY		844	GAGGCGCTTTCGTTTGTATTTGTGAGAGATTCAGCACTGAGAAAGAAAGCTGATA	903
Dd		276	CAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	217
OY		904	ATGCAAGAGAGTAAGAATTCGATACAAATGAAGCAAAAAAATAAAGAGCTTTGAAAT	963
Dd		216	ATACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	157
OY		964	TGAGTTTGTGTTTGTATAGGATTTTGTGAGTGAATAATAGAAATTCATGGAATGATATT	1022
Dd		156	AAAAATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	99
OY		1024	CTCACAAATAAATAAGACTCTGCAACTAATAAGAAATCAATGAAGAGGACACAA	1080
Dd		98	ATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	42

RESULT 2  
US-10-239-676-198/c  
; Sequence 198, Application US/10239676  
; Publication No. US20030082609A1

```

: GENERAL INFORMATION:
: APPLICANT: OLER, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
: FILE REFERENCE: 5013.1003
: CURRENT APPLICATION NUMBER: US/10/239,676
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: PCT/EP01/03968
: DE 10019058.8
: DE 10019173.8
: DE 10032529.7
: DE 10043826.1
: PRIOR FILING DATE: 2001-04-06
: 2000-04-06
: 2000-04-07
: 2000-06-30
: 2000-09-01
: NUMBER OF SEQ ID NOS: 228
: SEQ ID NO 198
: LENGTH: 7823
: TYPE: DNA
: ORGANISM: Artificial Sequence
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-198

Query Match          5.0%; Score 83.6; DB 14; Length 7823;
Best local similarity 44.6%; Pred. NO.3.3e-05;
Matches 474; Conservative 0; Mismatches 569; Indels 19; Gaps 3

```

Oy	8	CACAGCGAAATTTGAAAAAATGATAAAATTTTGGCAGGTATTAAAGCCGCTTAACA	67
Db	1085	CAAAAATNTNTATTATAAATATATATAATTAACCTATATTCTATAAACAAAAAAC	1028
Oy	68	AATGAGTAGACAGAGAAAAAATAATATATCATATTTTGAGTTAGTAAAAAGAGAA	127
Db	1025	CTTAACCTATCTATAAATAATATATTAATATATATTAATTAATCAATATATAAAAAAT	966
Oy	128	AAAAATGAAGAAGACCCTCGGCCTTAACAGTGGAAAAACGAGAAATATATAAAGAG	187
Db	965	TAAATTTTTACTTAACCTTAATAATTAACAAAAAAAAACAACAAAAATTAACCTTAA	906
Oy	188	ACTGTGATTTTAAATGCAATCGTAGAGAAAAAATTTTAATTTTCATTTTCGAGGA	247
Db	905	CATAAATACT-----ATAATTCAAAAAAAACAACGCAACATCATTTTCTTAATT	852
Oy	248	TTAATTTGGTTAGGTGATGAAAAATCTAGATNAANAATGCAGATCAAAATGCTTGA	307
Db	851	AAATATATCTACAACTTCTTAATTAACCAAATTAATTTACMAAAAAAATACGTACT	792
Oy	308	ATTGACATTTTGAATAGCTAGTANTATCAATTAATGGGGCTTGTCTAATTTATTTGC	367
Db	791	TTTCTTTTATTAACAACACTCACCAATTAATTAATCAATATATTAATTTCTTAAATAC	732
Oy	368	GAAATGGAANAATCGAGTGAAGAAAAATAGTTGGCGAGCAAAAAACCTGCCGTT	427
Db	731	TTAATAATTTAACACATCTCAAAATTTCTCAAAAAACAAAAACTACATCTTTAAAA	672
Oy	428	TTTTTCAATGACTTTGGAAAAAATTCATTTGTGAGCGGTAGCGAAACTTGAATTTTT	487
Db	671	CTTTACCCATCTATATATAATCTCACAAAAAATAATTAACACATTCATTAATTCAAA	612
Oy	488	ACATGGAAATTTGAAAAAATTAAGCGCAAAAGAACTCAATATGAAAAAATTTATTATA	547
Db	611	ATATTTTAAATTTAAAAAACTTAAACCTAGCAAAATACATACATTAATTTCTCAATAA	552
Oy	548	AAAAAGAGATCGGATATGATTTTAAAGCAGAAACTGC-----ATTGAATGAAGA	601
Db	551	CTAAACCAAAATTAACATAAAAAATCAAAACATPAATACTCTATTAACCAACAAAT	492
Oy	602	AAAAGATTTGAAAAAATCTATGCTGAGAGTGAATTAAGCAAAAAAATTTGGAGCTCA	661

Query Match	Similarity	4.5%	Score	7.2	DB	14	Length	18997
Best Local	Similarity	44.0%	Pred.	No.	0.0017			
Matches	498	Conservative	0	Mismatches	62	Indels	10	Gaps
QY	7	ACACAGGCCAAATTCGAAAAATAGATATAATTTTCGACGATTTAAACCGACATTTAAAC	66					
Db	9867	ATACAAAATTAATAAAAAATATCAAAAAAAATAAACCTCATCTACTATAAAAAATCAAAAAAA	9808					
QY	67	AAATGAGCGAAGAGAAAAGAAAAAAATTAATACATATTGTGAGTTAGTAAAGAGAAG	126					
Db	9807	ATTCGAAAAAACAAATTAACCAACCTTAATCTATATACCTTAATCCTTAACCTACTCAAAA	9748					
QY	127	AAAAAATTAAGAAGACCTCGGCTTAACAGTCGAAAAACCGAAATATAATAAAGAAAGA	186					
Db	9747	AACTTAACCAAAACACTCTTAACCCAAAAATTTAAACCAACCTTAACCAACATTAACAAA	96888					
QY	187	GACGTGATTTTAAATGGAATTCGTGAGAAAAGAAAAATTTTAATTTTCATTTTCGAGG	246					
Db	9687	CCTTAATCTCTTTAAAAACATTAACAAAAAAATTAATACCTTAACCTTAACCTTAACAAAT	96288					
QY	247	ATTAATTTGTTGATAGTGAATGAAAAATCTGATATAAAAATGCAGATCAAAAAATGTCTTG	306					

Db	9627	CTAAAAAATAAAAAATTTTAAAAACATCAAAACAAAAACAAATTAATTAACATATAA	9568
OY	307	AATTGGACATTATTGAATACGTAGTATATCAATATATGGGGTTTGTCATTTATTTTG	366
Db	9567	ATATTTCAAAAATTAACAATTAACCTCAATACCTTTTAAAAACAAATATATAAAAACAAAA	9508
OY	367	CGAAGATGAAAAATCTG---AGTGAAGAATAATAGTTGGAGAGCAAAAAACCCCTG	422
Db	9507	CAAAAAATAAAAACGTCAAAACACGAAAAAACCCAAATCATAAAAAACCTATATATCAAA	9448
OY	423	CCGTTTTCCTCAATGACCTTGGAAAAAATTCATTGGAGGGGTAGCAAACTTTGAAAT	482
Db	9447	AAAAAATTTTAAATTTTACTTTAAAAAACCAAAATTTATCATTAATAATTAACCATTCAAAT	9388
OY	483	TTTTTA--CATTGGAAATTTGAAAAAATAAGCAAAAGAAACCTAATGAAAAAATATT	540
Db	9387	TTACAAATCAAAATCAAACTCAAAACACAAATAAAAAATTAACCTAAAAACCGAA	9328
OY	541	ATTATAAAAAAGAGATCGATATGATTTTAAAAACAGAAAAACGTGACATTTGATA	600
Db	9327	AAACCAATAAAAAATTTACATATATCTTCATAAAAAATATTAATAAAACCTAAAAACAAA	9268
OY	601	AAAAAGATTTGGAAAAATCTATGCTGAGATGAAATTAAGCAAAAAATTTGGAACTC	660
Db	9267	ATAAAAAATATACATATATTAATAAAATATTTAAAAACAAACATATCAACACATTAATA	9208
OY	661	AACCCGGTGTGTTTATGAAAAATGACGATGAAGAAGAAATGATGAAAAATATCAACCTGATG	720
Db	9207	ACAAACCTCTAATCTCTGCAATTAACATTTATAAAAAATATATTAATAATCCAAATCTA	9148
OY	721	TTAATGAAGAAACAGCAGGTCAATATAGGAATTTATCAAAATTAAGTTAGACATAGTA	780
Db	9147	ATTAATATATTCAAAAAATACTTAACCCCAATCCACACMACCTA-ATPACTCCGTA	9089
OY	781	AATGAGATGATCTGTAAACGGGACTATTAGAGTGGACGAAATAGTTTGATATAA	840
Db	9088	AAACAAAAAATCTTAACCTAACTAACTCCTTACCTTACAAATTTCAACATTTTAAATTTTA	9029
OY	841	CAGAGAGTGCCTTTCGTTTGTATTTGTGAGAGAAATTCAGCAACTGAGAAAGAACCTG	900
Db	9028	TAAACTAAAAAACAACAACTAATTTTAAACAAATTTACCTCATATTAATTAATAAAC	8969
OY	901	ATAATGCAGAGCTAAAGATTTCCGTACATGAAGCAAAACCTTAAGAGGCTTTTG	960
Db	8968	AAAACAATAAAAATTAATAAATTAATTAACCTTATTTCAAAACCAAAACCTTATTTT	8909
OY	961	AATTGAGTGT---GTTTGTGATAGAGATTTTTCAGGCAAAATAGAAATCAATGAAATG	1017
Db	8908	AAAACAATCTTAATATTTTCTACTTATTCCTCTTAATTAACCTTTAATTAACACACT	8849
OY	1018	ATATTTTCACACAAAAAAGACCTCTGCACTAAATAAGAAAAACCAATGAAGAACCGGACA	1077
Db	8848	TTCAATTAAACAAAAAATAAAACCTTAATAAATTTACCAACACTTAATTAATTAACCTTA	8789
OY	1078	CAATGATGATATTTTACAGGCTAAAAAATAATTAATCTACATATGATGTC	1128
Db	8788	AAAACAATATCTCTTCTTAATAAATATCTAACTCAAAATTAACAAATCAAC	8738
RESULT 4			
US-10-239-676-51/c			
; Sequence 51, Application US/10239676			
; Publication No. US20030082609A1			
; GENERAL INFORMATION:			
; APPLICANT: OLEK, Alexander			
; APPLICANT: PIERPENROCK, Christian			
; APPLICANT: BERLIN, Kurt			
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation			
; FILE REFERENCE: 5013.1003			
; CURRENT APPLICATION NUMBER: US/10/239, 676			
; CURRENT FILING DATE: 2002-09-24			
; PRIOR APPLICATION NUMBER: PCT/EP01/03968			
; DE 10019058.8			

DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 51  
LENGTH: 9539  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-51

Query Match 4.3%; Score 71; DB 14; Length 9539;  
Best Local Similarity 43.6%; Pred. No. 0.0052;  
Matches 469; Conservative 0; Mismatches 595; Indels 11; Gaps 3;

QY 7 ACACAGGCAATTGAAAAATGATTAATTTTCGCGATTTAAAGCCGACTTAAC 66  
DB 6229 ATAAAAATTAATTTAAAAATCAATATTAATTAATCAAAAAATTTAAATTAAC 6170  
QY 67 AAATGAGTGAAGAGAAAAAATTAATTAATTTGAGTTTAAAGAGAAAG 126  
DB 6169 ATAAATTTTAAAAAACAATTAATTAATTAATTAATTAATTAATTAAT 6110  
QY 127 AAAAAATTAAGAAAGCTCGCTTAACAGTCGAAAAACCAATTAATTAAGAAAGA 186  
DB 6109 TCAAAACACATTTTAACTCAAAATTAATTAATTAATTAATTAATTAATTAAT 6050  
QY 187 GACTGTGATTTTAAATGGAATCGAGGAAAGAAAT-TTAAATTTTCAATTTTCGAGG 245  
DB 6049 AACGAAATTTAAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 5990  
QY 246 GATTAAATTTGTTAGTTGATGAATTAATTAATTAATTAATTAATTAATTAAT 305  
DB 5989 TCATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5930  
QY 306 GAATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 365  
DB 5929 AAAAAATCTATTTAAACCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 5870  
QY 366 GCGAAGATTTGAAATCTGAGTGAAGAAATAGTTTGGAGAGCAAAACCCCTTGGCG 425  
DB 5869 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5810  
QY 426 TTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 485  
DB 5809 TAAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5750  
QY 486 TTACATTTGGAATTTGAAAAATTAAGCAAAAGAAACCAATTAATTAATTAATTAAT 545  
DB 5749 TAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5691  
QY 546 AAAAAAGAGATGATGATGATTTTAAAGCAGAAATGATGATGATGATGATGATGAT 605  
DB 5690 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5631  
QY 606 GATTGGAATAATCTATGCTGAGAGTGAATTAAGCAAAAGAAATTTGGAGACTCAACC 665  
DB 5630 AAAAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5571  
QY 666 GGTGTGTTTGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 725  
DB 5570 AAAAACTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5511  
QY 726 GAAGAAGCAGAGCTCAATATAGAAATTTATCAAAATTAATTAATTAATTAATTAATTAAT 785  
DB 5510 ACATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5451

QY 786 GATGATCTAGTAACGGGACTATTAAGTGTGAACTCGAAATAGTTTGAATTAACAGCA 845  
DB 5450 AATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5391  
QY 846 AGTGCCCTTTCCTTTTGTGATTTTGTGAGAAATTCAGCACTGAGAAAGAAAGCTGATAT 905  
DB 5390 -----TTAATTTTAAAAAACAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5340  
QY 906 GCAAGAGATTAAGATTTGATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 965  
DB 5339 TCCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5280  
QY 966 AGTTTGTGTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1025  
DB 5279 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5220  
QY 1026 CACAACAAAAAGCTCTCAAGTAAAGAAAAACAATTAATTAATTAATTAATTAATTAATTAAT 1080  
DB 5219 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5165

RESULT 5  
US-09-960-352-14521

Sequence 14521, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengping

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960.352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 14521

LENGTH: 411

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 62-LIB3058-002-Q1-K1-H6

US-09-960-352-14521

Query Match 4.2%; Score 70; DB 11; Length 411;  
Best Local Similarity 54.7%; Pred. No. 0.0037;  
Matches 139; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 62 AAAAAATGAGTGAAGAGAAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 121  
DB 117 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 176  
QY 122 GAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 181  
DB 177 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 236  
QY 242 GAGGATTAATTTGTTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301  
DB 297 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 356  
QY 302 TGTGATTTGACA 315  
DB 357 TAAAAATTTTAA 370

RESULT 6

US-09-960-352-3400/C

Sequence 3400, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:



Query Match	4.28;	Score 69.4;	DB 11;	Length 446;
Best Local Similarity	48.48;	Pred. No. 0.0047;		
Matches 193; Conservative	0;	Mismatches 206;	Indels 0;	Gaps 0

RESULT 7  
US-09-878-574-4653/C  
; Sequence 4653, Application US/09878574

```

1 APPLICANT: Byrum, Joseph R.
2 APPLICANT: La Rosa, Thomas J.
3 APPLICANT: Thompson, Michael D.
4 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
5 TITLE OF INVENTION: Plants
6 FILE REFERENCE: 38-21(15401)B
7 CURRENT APPLICATION NUMBER: US/09/878,574
8 CURRENT FILING DATE: 2001-12-21
9 PRIOR APPLICATION NUMBER: 09/333,535
10 PRIOR FILING DATE: 1999-06-14
11 NUMBER OF SEQ ID NOS: 15775
12 SEQ ID NO 4653
13 LENGTH: 821
14 TYPE: DNA
15 ORGANISM: Glycine max
16 FEATURE:
17 NAME/KEY: unsure
18 LOCATION: (1)..(821)

```

Query Match	4.2%	Score 69	DB 11	Length 821
Best Local Similarity	44.3%	Pred No	0.0064	
Matches 270	Conservative	0	Mismatches 340	Indels 0
			Gaps	0

[illegible]

```

RESULT 8
US-10-198-846-6381/c
: Sequence 6381, Application US/10198846
: Publication No. US2003009997A1
: GENERAL INFORMATION:
: APPLICANT: Lillie, James
: APPLICANT: Xu, Yongyao
: APPLICANT: Wang, Youzhen
: APPLICANT: Steilmann, Kathleen
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
: TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF BREAST CANCER
: FILE REFERENCE: MRI-049
: CURRENT APPLICATION NUMBER: US/10/198,846
: CURRENT FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/306,220
: PRIOR FILING DATE: 2001-07-18
: NUMBER OF SEQ ID NOS: 14084
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6381

```

```

: LENGTH: 960
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
: LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
: LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
: LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427
: OTHER INFORMATION: n = A,T,C or G
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 428, 433, 450, 465, 466, 468, 471, 487, 510, 512, 513, 516,
: LOCATION: 517, 518, 528, 530, 534, 539, 553, 555, 559, 575, 580, 588,
: LOCATION: 590, 610, 625, 626, 642, 645, 656, 658, 659, 675, 685, 686,
: LOCATION: 690, 700, 702, 706, 716, 731, 745, 748, 756, 798, 802
: OTHER INFORMATION: n = A,T,C or G
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 807, 808, 819, 820, 821, 830, 838, 840, 844, 845, 849, 861,
: LOCATION: 871, 874, 875, 891, 892, 895, 898, 900, 901, 902, 904, 907,
: LOCATION: 917, 923, 925, 926, 927, 929, 955
: OTHER INFORMATION: n = A,T,C or G
: US-10-198-846-6381

```

```

Query Match 4.2%; Score 69; DB 14; Length 960;
Best Local Similarity 39.8%; Pred. No. 0.0067;
Matches 317; Conservative 0; Mismatches 468; Indels 12; Gaps 2;

```

```

QY 22 AAAAAATGATTAATTTTCGACGATTAAGCCGACTTAACAATGAGTGAAG 81
DB 951 AAAAAAAAAAAAAAAAAATTTGTTANNANNACTTANACCTTTAAANNTNNNANTNNAAN 892
QY 82 AAAAAAATAATATCATATTTGAGTAGTAAAGAGAAGAAAAATAAAGAG 141
DB 891 NAAAAAATATTTAAANNTANAAATTTTNNAAAACTAATTTNNTAANNTAATNAAATTTT 832
QY 142 ACCTCGGCTTAACGCAAAAAACAGAAATATATAAAGAAAGAGACTGTATTTTAA 201
DB 831 TTTAATAAANNNNNAAAAATTAANNTTTTNTAANTTATTAACCAAAATTTTAAAA 772
QY 202 TCGAATCGCTGAGGAAAGAAATTTTAATTTTCATTTTCGAGGATTAATTTGTTAA 261
DB 771 AAAAAATTTTAAANNTAATATTAATTAATTTTAAATTAATTAATTAATTAATTAAT 712
QY 262 GTTATGAAAAATCTAGATTAATAAATGACATCAAAATGCTGAATTTGACATTAATG 321
DB 711 TTTAANCAANNTTTTNTTTTNTTNNAAAAAANNTTTTAATTAATTAATTAATTAAT 652
QY 322 AATATCGTAGTATCAATATGAGGCTTGTCTATTTTATTTTTCGGAAGATGAAATC 381
DB 651 TTTTNTNANATATTAATAAANNTTTTAAAAAATTTTAAATTAATTAATTAATTAAT 592
QY 382 TGAGTGAAGAAATAGTTTTCGAGAGCAAAAAACCCGCTTTTTCGAATGACT 441
DB 591 TTTNT-----TTTCTNAAAAAANAAAAATTTTNNAAANNTTTTAAACCT 541
QY 442 TTGAAAAAATTCATTTGAGCGGTAGCGAAACCTTTGAATTTTACATTTGAATTTG 501
DB 540 TTAATTAANANNAATTTTNTTNNNTNNAANATTAATAATATTTTNTTAACT 481
QY 502 AAAAAATTAAGCAAAAGAACTCAATGCAAAAAATATTTATTAATAAAGAGAGATCGG 561
DB 480 AAAAAAANANANNAATTAATTAATTTTNNAAAAAATTAATTTTANTTANNNANANA 421
QY 562 ATATGATTTTAAAGCAAAAGCAATGACATTAATGAATAAATAATTTGAA---AAA 618
DB 420 AATTAATAANNTAANAAAAAATTTTANTTAATTAATAATTAATTAATTAATTAAT 361
QY 619 TCTATGCTGAGAGATTAATAAGCAAAAAAATTTGGAACCTCAACCCGCTGTTGTTAG 678
DB 360 TTTAATTTTAAATTAATAAATAAATTTTNTTATATTAATAAATAAATTTTATTAATAA 301

```

```

QY 679 AATGACGATGAAGAAATGATGAAAAATATCAACCTCGATTTAATGACAACAGCAG 738
DB 300 AAAAAAANNTAATTAATTAATAAAGNNNNNTNNTTAAAAAATAATTAATTAATTAATTA 241
QY 739 GTCAATATGAAATTTATTCAAAAATTAAGTGAATGATCAATCAAGATGATCTAGTAA 798
DB 240 ATTTAATTTGNAAAAAAATTAATTAATAAATTTATTTTATTAACANATGTGTATTAATA 181
QY 799 CGGACTATTTAGAGTGT 815
DB 180 NNTTAANAATAATGANT 164

```

```

RESULT 9
US-09-960-352-4630
: Sequence 4630; Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengbing
: APPLICANT: Byatt, John C.
: APPLICANT: Mathalagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 4630
: LENGTH: 327
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 20-LIB3058-048-Q1-K1-E11
: US-09-960-352-4630

```

```

Query Match 4.1%; Score 67.4; DB 11; Length 327;
Best Local Similarity 51.9%; Pred. No. 0.0097;
Matches 152; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```

```

QY 17 AATGAAAAATAGATTAATTTTCGACGATTTAAGCCGACTTAACAATGAGTGA 76
DB 32 AAGTAGAATTTTCATGAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 91
QY 77 AGAAGAAAAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 136
DB 92 AAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAA 151
QY 137 AGAAGACCTCGCTTAACAGCTGCAAAACCAAGAAATTAATAAAGAAAGAGCTGTGANT 196
DB 152 AAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAA 211
QY 197 TTTAATGAAATCGTGAAGAAAGAAATTTTAAATTTTCAATTTTCGAGGATTAATTTGT 256
DB 212 AGCGAGAGAAAGTTAAAGGAAAAAATTAATAAAGATTAATTTAAAGSTAAATTTT 271
QY 257 TGTAAAGTGAGCAAAATCTAGATTAATAAATGCAATCAAAAATGTGTTGAAT 309
DB 272 TAAAGGAAAAAATTAATAAAGGAGGGAATTTTAAAAAATTTTGAAT 324

```

```

RESULT 10
US-10-239-676-185/c
: Sequence 185; Application US/10239676
: Publication No. US20030082609A1
: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kutz
: TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
: FILE REFERENCE: 5013.1003
: CURRENT APPLICATION NUMBER: US/10/239,676
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: PCT/EP01/03968

```

```
DE 10019058.8
DE 10019173.8
DE 10033529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 185
LENGTH: 7657
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-185
```

```
Query Match
Best Local Similarity 47.0%; Score 67.4; DB 14; Length 7657;
Matches 209; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
```

```
QY 494 GAAATTTGAAAAATAGCGAAAAAGAACTCAATGGAATAATATTATTATTAATAAAG 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5979 GAACATTAAATATATCACCACAAAAAAGTAACGAAATATAAAAAATATAAATAA 5920
QY 554 GAGATCGATGATGATTTTAAAGCAGAAACTGACATTCATGAAAAAAGATTGGA 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5919 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 5860
QY 614 AAAAACTATGCTGAGAGTGAATTTAAAGCAAAAAAATTTGGACACCGGTGTGT 673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5859 TAATTAATAATATAAAAAAATATAAAAAAATATAAAAAAATATAAAAAA 5800
QY 674 TTGAATGAGCGATGAAAGAAATGATGAAAAATATCACTGATGTTAATGAGAAC 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5799 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 5740
QY 734 AGCAGCTCAATATGAAATTTATCAAAAAATTAAGTGAACATGTAATCATGATCT 793
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5739 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 5680
QY 794 AGTAACGAGCTTTAGAGTGTGAACTGCAATAGTTTGTATTAACAAGAGTCCCTT 853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5679 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 5620
QY 854 TCGTTTTGTATTGTGAGAGAAATTCAGCAACTGAGAAAAAGAGTGAATGCAAGAG 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5619 TAAATTAATCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 5560
QY 914 AGTAAGATTCGATCAATGAAA 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5559 ATAAAAATTAATAAATTAACCAATA 5535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11  
US-09-960-352-573/C

Sequence 573, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 573  
LENGTH: 428  
TYPE: DNA

ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 03-LIB3058-031-Q1-K1-A11  
US-09-960-352-573

Query Match  
Best Local Similarity 52.3%; Score 66.6; DB 11; Length 428;  
Matches 147; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

```
QY 502 AAAAAATAGCGAAAAAGAACTCAATGGAATAATATTATTATTAATAAAGAGATCGG 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 331
QY 562 ATATGATTTTAAAGCAGAAAACTGACATGTAATGAAAAAAGATTGGAATAATCT 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 330 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 271
QY 622 ATGCTGAGATGATTTAAAGCAAAAAAATTTGGAACTGACCCGGTGTGTTTGA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 270 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 211
QY 682 TGACGATGAAAGAAATGATGAAAAATATCACTGATGTTAATGAAAGACGAGTC 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 151
QY 742 AATATGAAATTTATTCAAAAAATTAAGTTGAGCATGATAA 782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 AATATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 12

US-10-239-676-35/C  
Sequence 35, Application US/10239676  
Publication No. US20030082609A1

GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8  
DE 10019173.8  
DE 10033529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 35  
LENGTH: 12405  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (7895)  
US-10-239-676-35

Query Match  
Best Local Similarity 42.8%; Score 65.4; DB 14; Length 12405;  
Matches 458; Conservative 0; Mismatches 596; Indels 15; Gaps 2;

```
QY 63 AATCAATGATGCTGAGAGAGAAAGAAAAATATATATATTTAGTTAGTAAAGAG 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9523 AAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAAA 9464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 AAGAAAAAATTAAGAGACCTCGGCTT-----ACAGTCGAAAAACCAAAATTAATAA 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 02:19:36 ; Search time 2699.14 Seconds  
(without alignments)  
11082.453 Million cell updates/sec

Title: US-08-876-132-2  
Perfect score: 1847  
Sequence: 1 ACATCTCAACACAGTTTAA...AACACAGCAAAAGACATCT 1847

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_nam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.8	5.8	1101	17	CNS00EVL
2	101.2	5.5	1101	17	CNS00EVL
3	100.6	5.4	1135	17	CNS0336Q
4	97.6	5.3	1201	17	CNS0167M
5	97	5.3	974	17	CNS0017T
6	96.2	5.2	1029	17	CNS01ZGM

7	94.4	5.1	829	17	CNS03LUB	AL24932 Tetradon
8	93.6	5.1	1025	17	CNS01472	AL104216 Drosophill
9	93.4	5.1	1101	17	CNS003BD	AL064091 Drosophill
10	93.2	5.0	1101	17	CNS0039G	AL063921 Drosophill
11	93	5.0	660	17	BH183498	BH183498 023_L-07-
12	92.2	5.0	660	17	CNS070M	AL620449 t3 end of
13	92.2	5.0	1036	17	CNS03LWJ	AL250012 Tetradon
14	91.8	5.0	804	17	B12681	B12681 F27D1-Sp6.1
15	91.8	5.0	1885	10	BE420745	BE420745 HMM002.B0
16	91.6	5.0	934	13	BM415088	BM415088 OP20159.M
17	91.6	5.0	1101	17	CNS0039R	AL063932 Drosophill
18	91	4.9	1101	17	CNS014XJ	AL104737 Drosophill
19	91	4.9	1187	17	B11102	B11102 F19C22-T7 I
20	90.4	4.9	942	17	CNS0186S	AL109318 Drosophill
21	90.2	4.9	982	13	BM415348	BM415348 OP0422.M
22	89.8	4.9	960	17	AG031602	AG031602 Pan trogl
23	89.6	4.9	782	10	BB622182	BB622182 BB622182
24	89.6	4.9	1101	17	CNS0021J	AL061936 Drosophill
25	88.2	4.8	1101	17	CNS0005X	AL060843 Drosophill
26	88	4.8	1101	17	CNS00E07	AL069440 Drosophill
27	88	4.8	1225	17	CNS0161D	AL106171 Drosophill
28	87.8	4.8	879	17	CNS01JRG	AL167405 Anopheles
29	87.8	4.8	1139	17	AQ097537	AQ097537 HS_3153_A
30	87	4.7	714	17	AQ271175	AQ271175 nbxb0024E
31	87	4.7	910	13	BM415636	BM415636 OP20714.M
32	86.8	4.7	1101	17	CNS012JN	AL101645 Drosophill
33	86.6	4.7	668	9	AL514901	AL514901 AL514801
34	86.6	4.7	907	17	CNS021J4	AL116953 Tetradon
35	86	4.7	795	17	AG044101	AG044101 Pan trogl
36	85.6	4.6	1059	17	CNS0022B	AL097133 Drosophill
37	85.4	4.6	994	17	CNS04N0J	AL298972 Tetradon
38	85.2	4.6	990	17	CNS0060I	AL065624 Drosophill
39	85	4.6	1036	17	CNS00598	AL057797 Drosophill
40	84.2	4.6	890	17	AQ026918	AQ026918 CIT-HSP-2
41	84.2	4.6	915	12	BG520365	BG520365 PS2C09.Y
42	83.8	4.5	714	17	AQ050116	AQ050116 nbxb0003B
43	83.6	4.5	948	13	BM415494	BM415494 OP20570.M
44	83.6	4.5	1190	17	CNS020N7	AL206908 Tetradon
45	83.4	4.5	876	17	CNS0096I	AL053529 Drosophill

## ALIGNMENTS

RESULT 1  
CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999  
LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29B23 of RPCL-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL069706.1 GI:4949849  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Drosophila melanogaster.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage ;  
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr

COMMENT  
- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammosser in Pletier de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCL-98 and was constructed by partial





[illegible]

OY		948	AATTTTCGACGCAAAAATATTTTGGATTTTTTAAAGCGCATCTTGCTACGCCACTTT	1007
Dd		242	AAA	301
OY		1008	TGCCATATTTTAAACCTGCATCTCTTTATAAGTTTAATAGATATATCCGTTAGTTATAAA	1067
Dd		302	AAA	361
OY		1068	GTAATGTTAAACAAGATGAACAACTATATATTTAAATTCGTGAATTTATTTGACA	1127
Dd		362	AAA	421
OY		1128	GTGATTTATTTATTTAAGAGATATATCTATTAGCTTAAATATTAACATAAAAAAGAG	1187
Dd		422	AAA	481
OY		1188	TAAATATATGATTTGTATTTTAAAAAAAGCTTAGAAAAATGAATPAGACTTTATAAAA	1247
Dd		482	AAAAAAAAAAAAAAAAAAATTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATT	541
OY		1248	AGACGCGATATCAAAATCTTTCTTACATCTGTCATCTTCTTGATATAGATAAGCATT	1307
Dd		542	AAAAAAAAAAAAAAAAAAATTTTAAATTTAAAAAATAAAAAAAAAAAAAAAAAATTAATTA	601
OY		1308	AAATGCGTGAATGTGGCGATTTATTAACCTCAAATTTATCATTGATGAAGCTTTGA	1367
Dd		602	AAAAAAAAAAAAAAAAAAATTTAAAAAATAAAAAAAAAAAAAAAAAATTAATTAATTA	661
OY		1368	TCTTCTTGATGTGAGCACAAATTTTCGGCTTTCAATTAATTAATPACGACACGAAA	1427
Dd		662	AAAAATTAATTTTAAAAAATAAAAAAAAAAAAWTAAATPAAAAAATAAATATATATAA	721
OY		1428	TGAATATTTATCATCAGCTAAACCTGAA- AATGATTTTAAAAAAATACGCTTTTATTC	1486
Dd		722	AAAAAAAAAAAAAAAAAAATTTACATTTTAAAAAATAAAAAAAAAAAAWTCYAAACCAAT	781
OY		1487	ATTCGATCAATTTGGAGAATTTAATTAATTCGATGATGAGTACATACTACATCAAGAAATGA	1546
Dd		782	AAAAAAAAAAAAAAAAAAATTTAAAYATTAATTTAAAAAACAAAAACACACAYAAAAA	841
OY		1547	TTTAAAGGTTTGATTAATTTACATATGTGAGAAATTAACCATTTGAATTA---TAAATTAAG	1603
Dd		842	AAATTAATTAATTAATTAATAA	901
OY		1604	AGAAATATTTGATTTTCGAATTAACAAAAAGATATAAAGAAAGTCAGCTTTAAAAATATC	1663
Dd		902	AAAAACAAAAAATAAATAAATTAATTTACTTTAAACCAAAAAAATAAAAAAAAAAAAAA	961
OY		1664	TGCATTTGATGATGTTTAAAGATAGATAAAAAATTAATTTGATGAGAAGATTTGCACT	1723
Dd		962	AAAAAAAAAAAAAAAAAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATTA	1021
OY		1724	AAAGATATTTGAAATATTCGTATGAAAATCCATTTGATGATGATCAATGATGSCATTAAGAA	1783
Dd		1022	ATTAATAAAAAAAAAAAAAAAAAAAATTCAAATTAATTTCAAAAAAATAAATAAATAA	1081
OY		1784	ATCACAATTCGCAGAGAAATGCTATGCTTCCGATTTCAAAAAAACAAGCAA	1837
Dd		1082	AAAAAAAAAAAAAAAAAAAAAAAAAATTAATTAATTTCAATTTATTTAAAAAATACTAA	1135

RESULT 4  
 CNS0167M/c  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
 BAEN13M24 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL106396  
 VERSION AL106396.1 GI:5621701  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;





[illegible]

QY	917	AGAAGCAAAACAAAATAAATAACAAAGAAAATTTTCGACGCAAAAATTTTGGCAAT	976
Db	1007	AAAAAAAAADAAAADAAAADAAAAGCGTAAATTTAAAAATRTAAAGTCTATGTTT	948
QY	977	TTTTTAAAGCGACTACTTGTCTACCGCACTTTTGGCATTTTAAAAACCGACTATCTTTAT	1036
Db	947	TTTTTAAAGCGACTACTTGTCTACCGCACTTTTGGCATTTTAAAAACCGACTATCTTTAT	1036
QY	1037	AAGTTAATAGATATATCCGTTAGATTATAAAGTATGTTAAAAACGAGTAAACACATTAAC	1096
Db	887	GGGGRARADAAAAGGTGATGAACGAAATTTTCTTAAATGCTATAAATTAATTAATTAATTAAT	828
QY	1097	TTATATATTTATTTCTGATATATATATTTTACAGTGTATTTTAAATATTAAGAGATATAT	1156
Db	827	AAAAAGKAGAKRGAGAGAAAATGKGCGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	768
QY	1157	CTATTTAGCTAATATATACTAAAAAAGAGTAAATATATGATTTGATTTAAAAAAG	1216
Db	767	TTTGTGTGTTAAAAATATGTTGTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	708
QY	1217	CATTAGAAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1274
Db	707	TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	648
QY	1275	ATACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1334
Db	647	TTWCTTTGTTKKAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	588
QY	1335	AAACTCAATTTATCATTTGATGAAGTTTGTATCTTGTATGTTGAGCAATTTTCGG	1394
Db	587	MAAGCGGTGCGKAKGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	528
QY	1395	CTGGGCTTTCATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1452
Db	527	DAAAPADGMAAAARAKRTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	468
QY	1455	AAATGATTTTAAAAAATCTCGTTCTTATTCATTCGATTCGATTCGATTCGATTCGATTCGATTCG	1514
Db	467	TKAKWATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	408
QY	1515	CGATGAGTGTGTCATATACATCAAGAATGATTAAGGATTTGATTAATTAATTAATTAATTAAT	1574
Db	407	AAATGCTTAT	348
QY	1575	AGAATATACCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1633
Db	347	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	288
QY	1635	ATTAAGAAAGTCACTCTTAAATAATCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	1694
Db	287	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	228
QY	1695	AAATTTATGATGAAGAGATTTCGACATAAGATTTGAAATATCTGATTAAGAAATCTT	1754
Db	227	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	168
QY	1755	ATTGATGATCCAGATGATGATGAAGAATACACATTCGACGAGAAAT	1805
Db	167	NNNNNNNNGNANNNNNNANNNNNNANNNNNNANNNNNNANNNANAT	117
RESULT 9			
CNS003BD			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			

[illegible]

QY 1575 AGAATAACCATAT 1588  
 | : : : : :  
 Db 1070 AAAAAHMMWMCWT 1083  
  
 RESULT 10  
 CNS0039G/C 1101 bp DNA linear GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 DEFINITION BAC080810 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL063921.1 GI:4941778  
 VERSION AL063921  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1101)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see http://www.fruitfly.org/The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazuhiro Osoegawa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.  
 FEATURES  
 source  
 1. 1101  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /note="end : TET3"  
 BASE COUNT 201 a 64 c 131 g 202 t 503 others  
 ORIGIN  
 Query Match 5.0%; Score 93.2; DB 17; Length 1101;  
 Best Local Similarity 18.0%; Pred. No. 2.3e-06;  
 Matches 128; Conservative 310; Mismatches 273; Indels 1; Gaps 1;  
 QY 943 AGAAAAATTTTCGACGCAAAAAATTTTGAATTTTAAAGCGGATCTTGCTACCGC 1002  
 | : : : : :  
 Db 1099 AARWDDTWDTRKDDMTKWTWMDRADRRWAGDADRWADDDGAGTWTWMMW 1040  
 | : : : : :  
 QY 1003 ACTTGGCATATTTAAACCTGACTCTTATAGTATAGTATATCCGTAGATT 1062  
 | : : : : :  
 Db 1039 WWWWATWDTWMDKMMWMTAAKTDTATWMTWRPAMRADWAGRGKRRDAATDADAG 980  
 | : : : : :  
 QY 1063 ATAAAGTATGTTAAAAAGCAATAAACAATACTTATATCTGCAATTAATT 1122  
 | : : : : :  
 Db 979 RRDGRRKKKKDDKDDGDDKGGKKKAARAKATWMDWMDKMKWDGAKDRAD 920  
 | : : : : :  
 QY 1123 TGACAGCATATTATATATTAAGAGATATCTATTACCTTAATATAAATAA 1182  
 | : : : : :  
 Db 919 DDDGAGDDDDGKGDADDTGTGTDKDDKMDWDKAGTGWGDATWMAATDMMWGM 860  
 | : : : : :  
 QY 1183 AGAGTAATATATGATGTG-TATTTAAAAAAGCATTAGAAATGAATGAACATA 1241  
 | : : : : :

Db 859 ADADWMTWDAADDDWADRDMDAMWMDAMWAGARTARDRRDMDRAGKRGARRDR 800  
 QY 1242 TAAAAAGACGGTGATCAATCTTCTTCAATCTTGATCTTGATATAGATAA 1301  
 | : : : : :  
 Db 799 KRADKRRADDRDAATWTTTWTTTTTRDIDMKKKTITWIRMAADRDDDDDDRRDA 740  
 | : : : : :  
 QY 1302 AGCATTAATGCGTGAATGCGGATATTAACCTCAAAATTTATCATATGATAAG 1361  
 | : : : : :  
 Db 739 GTAGKWRRTWRKRRRRDRDMDADDTARDDRGRGGDAGKCKTKGRKRRDR 680  
 | : : : : :  
 QY 1362 TTTTATCTTCTTGATGTTTGACACAAATTTGGCGCTTGCTTCAATAAATAACAG 1421  
 | : : : : :  
 Db 679 ATWDTTDAWMDAAWMTTDTDTDDWDRDRRKGARRRRTTAATAAWDMWTWAMDAK 620  
 | : : : : :  
 QY 1422 AGCAATGATATTTATCATCAGCAAGTAAAGTAAATATTTAAATATCTGCTCT 1481  
 | : : : : :  
 Db 619 WDKTRADRRMDRMAADTDTDKARADRAKARAWARARADRRARADRRWTKGTTAT 560  
 | : : : : :  
 QY 1482 TATTCATTCGATCAATTTGGGAAGATTTAATACAGATGAGTACATCATCAGA 1541  
 | : : : : :  
 Db 559 WTTMAARAAMWAMWMTATTTATWTTTWTTTTWTTTTAAWMAWMTATWA 500  
 | : : : : :  
 QY 1542 AATGATTAAGCATATAGTATTAATACATATGAGAGAAATATGATATGAATATAAATA 1601  
 | : : : : :  
 Db 499 AWTAAWMAAAAAAATAATTTTTTTTTTTTWTAAWMTATWTTTWTTTTMAATTT 440  
 | : : : : :  
 QY 1602 AGAGATATATTTGATTTTCAATCAATAAAGATATAAAGAAAGTACGCTCT 1653  
 | : : : : :  
 Db 439 TTTWTTTWWAATTTATTTTWTTTTAAWMAWMTATWTTTATATATATTT 388  
 | : : : : :  
 RESULT 11  
 BH183498 660 bp DNA linear GSS 19-OCT-2001  
 LOCUS 023 L.07-rev SmbAC1 Schistosoma mansoni genomic clone 023L07 5',  
 DEFINITION DNA sequence.  
 ACCESSION BH183498  
 VERSION BH183498.1 GI:16288814  
 KEYWORDS GSS.  
 SOURCE Schistosoma mansoni.  
 ORGANISM Schistosoma mansoni.  
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigedidae; Schistosomatidae; Schistosomatidae; Schistosoma.  
 1 (bases 1 to 660)  
 REFERENCE Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams  
 D.,L., Johnston,D., Loverde,P.T. and Le Paslier,D.  
 Construction and characterization of a Schistosoma mansoni  
 bacterial artificial chromosome library  
 Genomics 65 (2), 87-94 (2000)  
 JOURNAL 20247247  
 MEDLINE  
 COMMENT Other\_GSSs: 023\_L.07-21  
 Contact: Pierce RJ  
 INSERM U 167  
 Institut Pasteur de Lille  
 1 rue du Professeur A. Calmette, 59019-Lille, France  
 Tel: (33) (0)3 20877783  
 Fax: (33) (0)3 20877888  
 Email: Raymond.Pierce@pasteur-lille.fr  
 CNS sequencing ID=DC00A023CF04BP1  
 Plate: 023 row: L column: 07  
 Seq primer: M13 reverse primer  
 Class: BAC ends  
 High quality sequence stop: 660.  
 Location/Qualifiers  
 1. 660  
 /organism="Schistosoma mansoni"  
 /strain="Puerco-Rican"  
 /db\_xref="taxon:6183"  
 /clone="023L07"  
 /clone\_lib="SmbAC1"  
 /sex="mixed"  
 /dev\_stage="cercariae"  
 /lab\_host="Biomphalaria glabrata"

/note="Vector: pBelobAC 11; Site 1: Hind III; Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pBelobAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold."

BASE COUNT 139 a 16 c 9 g 446 t 50 others

ORIGIN

Query Match 5.08; Score 93; DB 17; Length 660;  
Best Local Similarity 43.6%; Pred. No. 3e-06;  
Matches 278; Conservative 35; Mismatches 323; Indels 1; Gaps 1;

1061 TTATAAGTATGTTAAACGAGTAACAATTAATTAATTAATTCGATTAATA 1120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
640 TTTTAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1121 TTGACAGTATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1180  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
580 TATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 521  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1181 AAGAGGTAATATATGATGTTGTTTAAAGCATTAGAAATGAATGAACATT 1240  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
520 AATADTKAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 461  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1241 ATTAATAAGCGGTGATCAATCTCTGACATACATCTGATGATATGATTA 1300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
460 TAAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 401  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1301 AAGCATTAAATGCGTGAATGCGGATATTAATTAATTAATTAATTAATTAATTAAT 1360  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
400 AGGWTATATAAAGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 341  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1361 GTTTGATCTCTGTTGTTGAGCACAATTCGCGCTTCAATTAATTAATTAATTAAT 1420  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
340 TATATATATACAC-CTGACTCTCTDKAAAGGTAATCCCAAGAAATAAATAAATAAATA 282  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1421 GACGAATGATATTAATCATCATCAAGTAAGTAATGATTTAAATAATCACTGCTCT 1480  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
281 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 222  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1481 TTAATCATTCGATCAATTTGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1540  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
221 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 162  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1541 AATGATTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1600  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
161 AAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 102  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1601 AAGAGATATATGATTTGGAATTAACAAGATTAATAAAGAAAGTCAGCTCTTAATAAT 1660  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
101 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 42  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1661 ATCTGCATTCGATGTTTAAAGTAGATGAATA 1697  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
41 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5

RESULT 12  
CNS070NJ/c 660 bp DNA linear GSS 10-OCT-2001  
LOCUS T3 end of clone 023CF04 of library SmbAC1 from strain Puerto-Rican  
DEFINITION of Schistosoma mansoni, genomic survey sequence.  
ACCESSION AL620449  
VERSION AL620449.1 GI:16034591  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni.  
ORGANISM Schistosoma mansoni.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeolida; Schistosomatoidea; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 660)  
AUTHORS Le Paslier, M.C., Pierce, R.J., Merlin, F., Hiral, H., Wu, W.,

TITLE Williams, D.L., Johnston, D., Loyerde, P.T. and Le Paslier, D.  
JOURNAL Construction and characterization of a Schistosoma mansoni  
MEDLINE bacterial artificial chromosome library  
2024/247  
PUBMED 10783255  
REFERENCE 2 (bases 1 to 660)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage;  
BP 191 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr  
- Web: www.genoscope.cns.fr)  
COMMENT Partially Hind III digested and size-selected S. mansoni cercarial  
DNA was ligated into Hind III digested pBelobAC 11 vector and used  
to transform E. coli DH10B. The complete library contains 23808  
clones from 4 independent sizing-ligation-transformations. Average  
insert size ranges from 70-127 kb and genome coverage is 7.9-fold.

FEATURES  
source  
1. 660  
/organism="Schistosoma mansoni"  
/strain="Puerto-Rican"  
/db\_xref="taxon:6183"  
/clone="023CF04"  
/clone\_lib="SmbAC1"  
/note="end : T3"

BASE COUNT 139 a 16 c 9 g 446 t 50 others

ORIGIN

Query Match 5.08; Score 93; DB 17; Length 660;  
Best Local Similarity 43.6%; Pred. No. 3e-06;  
Matches 278; Conservative 35; Mismatches 323; Indels 1; Gaps 1;

1061 TTATAAGTATGTTAAACGAGTAACAATTAATTAATTAATTCGATTAATA 1120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
640 TTTTAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1121 TTGACAGTATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1180  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
580 TATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 521  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1181 AAGAGGTAATATATGATGTTGTTTAAAGCATTAGAAATGAATGAACATT 1240  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
520 AATADTKAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 461  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1241 ATTAATAAGCGGTGATCAATCTCTGACATACATCTGATGATATGATTA 1300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
460 TAAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 401  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1301 AAGCATTAAATGCGTGAATGCGGATATTAATTAATTAATTAATTAATTAATTAATTAAT 1360  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
400 AGGWTATATAAAGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 341  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1361 GTTTGATCTCTGTTGTTGAGCACAATTCGCGCTTCAATTAATTAATTAATTAATTAAT 1420  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
340 TATATATATACAC-CTGACTCTCTDKAAAGGTAATCCCAAGAAATAAATAAATAAATA 282  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1421 GACGAATGATATTAATCATCATCAAGTAAGTAATGATTTAAATAATCACTGCTCT 1480  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
281 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 222  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1481 TTAATCATTCGATCAATTTGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1540  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
221 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 162  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1541 AATGATTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1600  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
161 AAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 102  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1601 AAGAGATATATGATTTGGAATTAACAAGATTAATAAAGAAAGTCAGCTCTTAATAAT 1660  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
101 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 42  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1661 ATCTGCATTCGATGTTTAAAGTAGATGAATA 1697  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||





Fax: 215-898-8780  
Email: jecker@atgenome.bio.upenn.edu

Seq primer: Sp6  
Class: BAC ends  
High quality sequence start: 335  
High quality sequence stop: 346.  
Location/Qualifiers

## FEATURES

source

1..804

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="F27D1"

/clone\_lib="IGF"

/sex="hermaphrodite"

/note="Vector: BelobAC11; Site\_1: EcoRI; Site\_2: EcoRI;

Produced by Thomas Altmann"

BASE COUNT 241 a 14 c 20 g 439 t 90 others

ORIGIN

Query Match 5.0%; Score 91.8; DB 17; Length 804;

Best Local Similarity 41.6%; Pred. No. 4.5e-06;

Matches 333; Conservative 0; Mismatches 453; Indels 1; Gaps 1;

1012 ATATTTAAACCTGATCTTTATATAGTATATATCCGTAGATTATTAAGTAT 1071  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
788 ATAAAT 729  
1072 GTTAAAAACGAGTAAAAACATTAATCTATATATCTGATATATTTGACGTGA 1131  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
728 ATTAATAAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 669  
1132 TTATTTAATATATATAGATATATATATATATATATATATATATATATAT 1191  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
668 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 609  
1192 TATATGATGTGTATTTAAAAAGCATTAGAAATGAATAGACATTATAAAAAG 1251  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
608 ATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 549  
1252 GGTGATATCAATCTTCTTCAATCTTGCATCTTGTGATATAGATTAAGATTAA 1311  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
548 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 489  
1312 GGTGATGATGTGGCATATTTATTAATCAATTTTCAATGATGAAGTTTGAATCT 1371  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
488 AATTTAAAAAATTTTATTTTNNAAAAATTTTAAATTTTNTTAAAT 429  
1372 CTGTGATGTCAGCACAAATTCGGCTTTCATTAATAATATACAGACGAATGA 1431  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
428 -ANNNTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 370  
1432 TATTTTCATCAGCTAAACTGAAATGATTTTAAAAAATACGCTCTTTATTCATCG 1491  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
369 TTTTCTTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 310  
1492 ATCAATTTGGAGAAATTTATATGATGATGATGATGATGATGATGATGATGAT 1551  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
309 AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 250  
1552 GGATTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1611  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
249 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 190  
1612 TTGATTTGCAATTAACAAAGATTAATTAATTAATTAATTAATTAATTAATTA 1671  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
189 AATTTTNTAANNNNAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 130  
1672 TCGATGTTTAAAGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1731  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
129 GTTACCCNN 70  
1732 TGAATAATCGATGAAAAATCCTATGATGATGATGATGATGATGATGATGATGAT 1788  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

Db 69 CTNCCNNNNACNTNTAAACTCCGAGAGCAAGCAAGATNNAATGANAANAACCC 13

## RESULT 15

BE420745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1885 bp mRNA linear EST 24-JUL-2000  
HMM002.802 ITEC HMM Barley Leaf Library Hordeum vulgare cDNA clone  
HMM002.802, mRNA sequence.  
BE420745  
BE420745.1 GI:9418588  
EST.  
Hordeum vulgare.  
Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;  
Triticeae; Hordeum.  
1 (bases 1 to 1885)  
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,  
S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,  
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,  
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,  
Pechioni, N., Qualset, C., Schuch, W., Selvaraj, G., Sharifou, M.,  
Sorrells, M., Wamburton, M. and Wenzel, G.  
International Triticeae EST Cooperative (ITEC): Production of  
Expressed Sequence Tags for Species of the Triticeae  
Unpublished (2000)  
Contact: Herrmann RG  
Botanisches Institut der LMU  
Menzinger Str. 67, D-80638 Muenchen GERMANY  
Fax: 49 30 171683  
Email: herrmann@botanik.biologie.uni-muenchen.de  
International Triticeae EST Cooperative (ITEC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers  
1..1885  
/organism="Hordeum vulgare"  
/cultivar="Barke"  
/db\_xref="taxon:4513"  
/clone="HMM002.802"  
/clone\_lib="ITEC HMM Barley Leaf Library"  
/tissue="leaf"  
/dev\_stage="14 day old"  
/note="Vector: pBluescriptSK(-); 850 bp average insert  
size."  
BASE COUNT 1138 a 219 c 212 g 176 t 140 others

ORIGIN

Query Match 5.0%; Score 91.8; DB 10; Length 1885;

Best Local Similarity 38.2%; Pred. No. 3.4e-06;

Matches 354; Conservative 0; Mismatches 572; Indels 0; Gaps 0;

919 AAGCAAAACAAAAAATTAACAAAGAAAAATTTTGAGGAGAAAAATATTTGGAAATT 978  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
746 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 805  
979 TTTAAGGCGATCTGCTACCGCAGCTTTTGCATATTTTAAACCTGACTATCTTATTA 1038  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
806 AANGGGGGGCGCCGCCCTTCCTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 865  
1039 GTTAATAGATATATCGTTAGATTATTAAGTATGTTAAAAACGATAAAAACATATT 1098  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
866 CTTCCTTCANNGCTTTTNAAGAAAAAATTAATTAATTAATTAATTAATTAATTAATTA 925  
1099 ATATATTTAATTTGCAATTTATTTGACAGTGTATTATTAATATATTAAGATATATCT 1158  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
926 NNN 985  
1159 ATTAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1218  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
986 NAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1045  
1219 TTGAAAAATGAAATAGAACATTTTAAAAAAGAGCGTGTATCAATCTTCTTACATATC 1278  
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

```
Db 1046 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1105
QY 1279 TTGCATTACTTGTATAGATTAAGCATTAAATGGTGAATGCGCATATTATAAC 1338
Db 1106 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1165
QY 1339 TCAATTTCATTCATGGAAGTTTGTCTTGTGATGTTGACACAAATTCGGCTGG 1398
Db 1166 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1225
QY 1399 GCTTCATTAATAATATCAGAGCAATGAATATTATTCATCAGCTAAACCTGAAAT 1458
Db 1226 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1285
QY 1459 GATTTTAAAAAATCTCTCTTTTATTCATTCGATCAATGGAGAAATTTAATTACGAT 1518
Db 1286 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1345
QY 1519 GAGATGAGTACATACATCAGAAATGATTAAGATTAGATTAATTACACATATGGAGAA 1578
Db 1346 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1405
QY 1579 ATAACCATATGATATATAATAAGAAATATATTGATTTCGAATAACAAAGATATTA 1638
Db 1406 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1465
QY 1639 AAGAAAGTCAGCTCTTAAAAATATCTGCATTCGATGATGTTTAAAGTAGATGAAAAAT 1698
Db 1466 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1525
QY 1699 TTATTGATGAGAGGATTTGCACTAAAGATATTGAAAAATATCGTAGAAAAATCCTATTG 1758
Db 1526 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1585
QY 1759 ATGATCCAGATGATGCAATTAAGAAATCACAATTCGCAAGAAATGCTATGCTTCC 1818
Db 1586 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1645
QY 1819 GCATTAAAAAACAAGCAAAAAAGAGA 1844
Db 1646 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1671
```

Search completed: July 17, 2003, 06:38:58  
Job time : 2708.64 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 21:08:10 : Search time 377.725 Seconds

(Without alignments)  
9896.940 Million cell updates/sec

Title: US-08-876-132-1

Perfect score: 1660  
Sequence: 1 AGATCTACACAGCAAAAT.....AAGACTCGAGCAGATCT 1660

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1657	99.8	1660	AAV74275	Pantoea citrea CTY
2	109	6.6	6292	AA546735	Tumour suppressor
3	94.8	5.7	778	ABO15588	Oligonucleotide fo
4	94.8	5.7	778	ABO15589	Oligonucleotide fo
5	93.6	5.6	6644	AA333181	Base sequence of t
6	93.6	5.6	7372	AA333182	Base sequence of t
7	93.6	5.6	7797	AA333180	Compox virus bsr
8	93.6	5.6	7996	AA333184	Base sequence of t
9	90.2	5.4	14006	ABL33958	Human immune syste

C 10	89.8	5.4	6668	24	ABL33697	Human immune syste
C 11	88.2	5.3	975	24	ABO29508	Oligonucleotide fo
C 12	88.2	5.3	975	24	ABO29509	Oligonucleotide fo
C 13	87.6	5.3	7442	22	AA546686	Tumour suppressor
C 14	86.6	5.2	15548	24	ABL34155	Human immune syste
C 15	83.8	5.0	9539	22	AA545347	Chemically pretrea
C 16	83.8	5.0	9539	24	ABK28180	DNA transcription
C 17	83.6	5.0	7571	24	ABL32527	Human immune syste
C 18	83.6	5.0	7823	22	AA545490	Chemically pretrea
C 19	83.6	5.0	7823	24	ABL34061	Human immune syste
C 20	83.6	5.0	7823	24	ABK28418	DNA transcription
C 21	83.6	5.0	7823	24	ABK31493	Signal transductio
C 22	82	4.9	7814	22	AA546530	Tumour suppressor
C 23	81.6	4.9	11745	24	ABK28332	DNA transcription
C 24	80	4.8	7597	24	ABL33013	Human immune syste
C 25	80	4.8	12237	24	ABL34358	Human immune syste
C 26	79.8	4.8	5276	24	ABL32150	Human immune syste
C 27	79	4.8	6419	24	ABL32267	Human immune syste
C 28	78.4	4.7	2943	21	AA470229	Plasmodium falcipa
C 29	78.4	4.7	15674	24	ABL70514	Chemically treated
C 30	78.4	4.7	15674	24	ABL32363	Human immune syste
C 31	78.4	4.7	15674	24	ABL34477	Human immune syste
C 32	78.2	4.7	9964	24	ABL32098	Human immune syste
C 33	76.8	4.6	6314	24	ABL54314	Chemically treated
C 34	76.6	4.6	5641	24	ABL33396	Human immune syste
C 35	76.4	4.6	1200	24	ABO39210	Oligonucleotide fo
C 36	76.4	4.6	1200	24	ABO39211	Oligonucleotide fo
C 37	76.2	4.6	8079	24	ABL2313	Chemically treated
C 38	76	4.6	15387	24	ABL32184	Human immune syste
C 39	75.4	4.5	1998	21	AA470212	Plasmodium falcipa
C 40	75.2	4.5	6159	24	ABN80236	Human chemically m
C 41	75.2	4.5	7461	24	ABL33784	Human immune syste
C 42	75	4.5	1030	24	ABO42518	Oligonucleotide fo
C 43	75	4.5	1030	24	ABO42519	Oligonucleotide fo
C 44	74.2	4.5	18997	24	ABK33948	Human DNA for stag
C 45	74.2	4.5	18997	24	ABL32570	Human immune syste

#### ALIGNMENTS

RESULT 1	
ID	AAV74275 standard; DNA: 1660 BP.
AC	AAV74275;
DT	16-JUN-1999 (first entry)
XX	
XX	Pantoea citrea cryptic plasmid ps.
KW	Cryptic plasmid; ps; higher temperature; growth; elimination;
KW	mobilization; ds.
XX	
OS	Pantoea citrea.
XX	
FT	key
FT	Location/Qualifiers
FT	480..1595
FT	/*tag= a
FT	/note= "largest open reading frame"
XX	
XX	WO9859054-A1.
PD	30-DEC-1998.
XX	
XX	22-JUN-1998; 98WO-US12945.
XX	
PR	23-JUN-1997; 97US-0876132.
XX	
PA	(GENV ) GENENOR INT INC.
XX	
PI	Causey SC, Fowler T;
XX	

DR WPI; 1999-105625/09.  
DR P-PSDB; AAW90181.

PT Method for preparing an improved Enterobacteriaceae strain - useful  
PT for improving bacterial fermentation strains of the family  
PT Enterobacteriaceae

PS Claim 18; Fig 1A-C; 30pp; English.

CC The sequence is that of one half of the cryptic plasmid pS from *Pantoea*  
CC *citrea*. It can be used as part of a method of eliminating the cryptic  
CC plasmid from an *Enterobacteriaceae* strain which is useful for  
CC reducing the mobilization properties of plasmids residing within  
CC these strains. The elimination of the cryptic plasmid from the  
CC *Enterobacteriaceae* strain permits growth of the organism at a higher  
CC temperature, which decreases the production time for desired compounds  
CC in the carbohydrate pathway. It also has the commercial benefit of  
CC reducing both capital cost and starting materials cost of large-scale  
CC *Enterobacteriaceae* biocatalysis in the production of desired end  
CC products, e.g. 2-keto-L-gluconic acid (2-KLG), a precursor of ascorbic  
CC acid.

SQ Sequence 1660 BP; 663 A; 198 C; 340 G; 456 T; 3 other;

Query Match	99.8%	Score 1657;	DB 20;	Length 1660;
Best Local Similarity	100.0%	Pred. NC. 1.3e-285;		
Matches 1660; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	AGATCTACCAAGGCAAAATGAAAAATAGATAAAATTTTGGCAGATTTAAACCCAGT	60
Db	1	AGATCTACCAAGGCAAAATGAAAAATAGATAAAATTTTGGCAGATTTAAACCCAGT	60
QY	61	TAAACCAATGAGTGAAGAGAAAAAATAATATATATATTTGAGTTAGTTAAAG	120
Db	61	TAAACCAATGAGTGAAGAGAAAAAATAATATATATATTTGAGTTAGTTAAAG	120
QY	121	AGAAACAAAAATTAAGAGAGACCTCGGGCTTACGCTGAGAAAAACGAAATATAA	180
Db	121	AGAAACAAAAATTAAGAGAGACCTCGGGCTTACGCTGAGAAAAACGAAATATAA	180
QY	181	GAAAGAGACTGATTTTAAATGGAATCGTAGAGAAAAAATTTAAATTCATTTT	240
Db	181	GAAAGAGACTGATTTTAAATGGAATCGTAGAGAAAAAATTTAAATTCATTTT	240
QY	241	CGAGGATTAATTTGTTGTAAGTATGAAAAATCTAGATAAAAATGCGAGATCAAAAT	300
Db	241	CGAGGATTAATTTGTTGTAAGTATGAAAAATCTAGATAAAAATGCGAGATCAAAAT	300
QY	301	GTTGTAATTTGACATTAATTAAGAAATACGTAGATATCAATATGAGGGTGTCTATTT	360
Db	301	GTTGTAATTTGACATTAATTAAGAAATACGTAGATATCAATATGAGGGTGTCTATTT	360
QY	361	ATTTTCGGAAGATTGAAAACTGAGTGAAGAAAAATAGTTGGCAGAGCAAAAAACCT	420
Db	361	ATTTTCGGAAGATTGAAAACTGAGTGAAGAAAAATAGTTGGCAGAGCAAAAAACCT	420
QY	421	TGCGGTTTTTCAATGACTTTGGAAAAAATTCATTGTAGCGGTGCGCAAACTTTGAA	480
Db	421	TGCGGTTTTTCAATGACTTTGGAAAAAATTCATTGTAGCGGTGCGCAAACTTTGAA	480
QY	481	ATTTTTCACATTGGAATTTGAAAAAATTAAGCAAAAAAATCTCAATGAAAAAATTT	540
Db	481	ATTTTTCACATTGGAATTTGAAAAAATTAAGCAAAAAAATCTCAATGAAAAAATTT	540
QY	541	ATTATATAAAAAAGGAGATCGGATATGATTTTAAAAAGCAAAAACTGACATTTGAATGAA	600
Db	541	ATTATATAAAAAAGGAGATCGGATATGATTTTAAAAAGCAAAAACTGACATTTGAATGAA	600
QY	601	AAAAAGATTGGAAAAAATCTATGCTGAGAGGAAATTTAAAGCAAAAAAATTTGGGAATC	660
Db	601	AAAAAGATTGGAAAAAATCTATGCTGAGAGGAAATTTAAAGCAAAAAAATTTGGGAATC	660
QY	661	AAACCGGTGTTGTTTAGAATGACGATGAAGAAATGATAAAAATATCAACCTGCATG	720

Db	661	AACCCGGTGTGTTTGAAGATGACGATGGAAGAAATGATGAAAAATATCAACCTGCATG	720
OY	721	TTAAATGAGAANAACGACGAGTCAATATAGAAAATTATTCAAAAATAAAGTTGAGCATGTGA	780
Db	721	TTAATGAAAGAAACGACGAGTCAATATAGAAAATTATTCAAAAATAAAGTTGAGCATGTGA	780
OY	781	AATACAGATGATCTATGTAACGGGACTATTAGGTGTGGAACTCGGAATAGTTTTGATTA	840
Db	781	AATACAGATGATCTATGTAACGGGACTATTAGGTGTGGAACTCGGAATAGTTTTGATTA	840
OY	841	CAGAAGTGCCTTTCGTTTTTGATTTGTGAGAAATTCAGCAACTGAGAAAAGAACTG	900
Db	841	CAGAAGTGCCTTTCGTTTTTGATTTGTGAGAAATTCAGCAACTGAGAAAAGAACTG	900
OY	901	ATATGCAAGAAGGCTAAAGATTTTCGATACAAATGAAAAGCAAAAACCTAAAGAGCTTTTG	960
Db	901	ATATGCAAGAAGGCTAAAGATTTTCGATACAAATGAAAAGCAAAAACCTAAAGAGCTTTTG	960
OY	961	AATTGAGTTTGTGTTTGTATGATTAAGGATTTTTTGATGAAAATAGAAATTCATGCAATGATA	1020
Db	961	AATTGAGTTTGTGTTTGTATGATTAAGGATTTTTTGATGAAAATAGAAATTCATGCAATGATA	1020
OY	1021	TTTTCACACAAAAAAGACTCTGCAAGTAAAGAAAAACATGAAAGACGGACACAA	1080
Db	1021	TTTTCACACAAAAAAGACTCTGCAAGTAAAGAAAAACATGAAAGACGGACACAA	1080
OY	1081	TGATGATATTTTTTAAAGGCTAAATAATTAATCTACATATGATCGTTATCTGAT	1140
Db	1081	TGATGATATTTTTTAAAGGCTAAATAATTAATCTACATATGATCGTTATCTGAT	1140
OY	1141	TCCTTTCTATTTGTTGATTAACAGGTTGCAGACGACAGAGAATTTTAAAGGATATAGAGA	1200
Db	1141	TCCTTTCTATTTGTTGATTAACAGGTTGCAGACGACAGAGAATTTTAAAGGATATAGAGA	1200
OY	1201	TAGTAACAAAACAGATAGAGATGAGTATCTTTTAAATACCTGGTGGCAAAAGGTTGAA	1260
Db	1201	TAGTAACAAAACAGATAGAGATGAGTATCTTTTAAATACCTGGTGGCAAAAGGTTGAA	1260
OY	1261	ATGACAGAGGCAAGGCAAGAAAGACATTACATTTTGATTTATCAAAATATCATGATATAG	1320
Db	1261	ATGACAGAGGCAAGGCAAGAAAGACATTACATTTTGATTTATCAAAATATCATGATATAG	1320
OY	1321	AGCAAAATGATATATTTTGTCCGAATTAAGATTAATTTTTCTACAAACCAATG	1380
Db	1321	AGCAAAATGATATATTTTGTCCGAATTAAGATTAATTTTTCTACAAACCAATG	1380
OY	1381	GGAAGCTCTACAAAGCTTGAGCAATACCTCTCATATCCAAATAGAAAGCTTTCACTGT	1440
Db	1381	GGAAGCTCTACAAAGCTTGAGCAATACCTCTCATATCCAAATAGAAAGCTTTCACTGT	1440
OY	1441	ATACACTTTCGTACAGAGGTTGCGATGATCTCAAGCAATCCGCTGCGATGACTTCAACA	1500
Db	1441	ATACACTTTCGTACAGAGGTTGCGATGATCTCAAGCAATCCGCTGCGATGACTTCAACA	1500
OY	1501	TAGGCGCTNTTTTGGGTACAGATGTGACTCAAAACCGAGAGTTACTACGGCTATGCTGCT	1560
Db	1501	TAGGCGCTNTTTTGGGTACAGATGTGACTCAAAACCGAGAGTTACTACGGCTATGCTGCT	1560
OY	1561	TCGTCGNAAGGTGATCGCTGTAACTGATGTTGAGTGTCTGATGTTGTGNAAGCAAC	1620
Db	1561	TCGTCGNAAGGTGATCGCTGTAACTGATGTTGAGTGTCTGATGTTGTGNAAGCAAC	1620
OY	1621	AAGAGTCAGTTTNGCTGTATCAAGACATCCGACGCAATCT	1660
Db	1621	AAGAGTCAGTTTNGCTGTATCAAGACATCCGACGCAATCT	1660

RESULT	2
AAS46735/c	
ID	AAS46735 standard; DNA; 6292 BP
XX	
AC	AAS46735;



RESULT 3  
ABQ15588/C  
ID ABQ15588 standard; DNA: 778 BP.  
XX  
AC ABQ15588;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 2179.  
XX  
KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX Sequence 778 BP; 111 A; 24 C; 29 G; 614 T; 0 other;

Query Match 5.7%; Score 94.8; DB 24; Length 778;  
Best Local Similarity 46.4%; Pred. No. 4.7e-08;  
Matches 345; Conservative 0; Mismatches 397; Indels 2; Gaps 1;

QY 55 CCGACTTAACCAATGAGTGAAGAGAAAGAAAAATATACATATTTTGAGTTAG 114  
DB 776 CCGAAAAAATATACGAAAAAATTTACTATTAATAAAAAATATATAAAAA 717  
QY 115 TAAAAAGAGAAAAATTAAGAGAGCTCGCTTACAGCTGAAAAACAGAAATTA 174  
DB 716 AAATTACGAAAAAATTTACGAAAAAATTTATTAATAAAAAAATTA 657

QY 175 TAAAAAGAGAGAGCTGTGATTTTATGAAATCGTAGAGAAAAAGAAATTTATTTT 234  
DB 656 TAAAAAATTTTACGAAAAAATATATATTAATAAAAAAATTTTAAAAA 597  
QY 235 CATTTCGAGGATTAATTTGTTGTAAGTTGATGAAAAATCTAGATTAATAAATGAGATC 294  
DB 596 AAAATTACGAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTT 537  
QY 295 AAAAATGTGTGAATTTGACATTTATTAATACGATATATTCATTAATGGGGTTGTC 354  
DB 536 ACGAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTTTACGAA 477  
QY 355 TATTTTATTTTTCGAGAGATGAAATTCGAGTGAAGAAATAGTTTCGAGAGCAAAA 414  
DB 476 AAAAAAATTTACGAAAAAATTTATTAATAAAAAAATTTACGAAAA--AAAA 419  
QY 415 AACCTTCGCGTTTTCCTCAATGACTTTGCAAAAAATTCATTTGACGCGTAGCAAAC 474  
DB 418 AAAAATTTCAAAAAACGAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTTGCG 359  
QY 475 TTTGAATTTTTCATTTGAAATTTGAAAAATTAAGCAAAAGCAAACTCAAAATGAGAAA 534  
DB 358 AAAAAAATTTTACGAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTTACGAAAA 299  
QY 535 AATATTTATTAATAAAAAAGAGATCGATATGATTTTAAAGCAGAAACTGACATTGA 594  
DB 298 AAAATTACGAAAAAATTTATTAATAAAAAAATTTACGAAAAAATTTATTCGAAA 239  
QY 595 ATGAAAAAATGATTTGCAAAAAATCTATGCTGAGAGATTAATAACCAAAAAATTTG 654  
DB 238 AAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTTATTCGAAAAA 179  
QY 655 GAATCAACCGGTGTTGTTTACGAATGACGATGAAGAAATGATGAATAATATCAAC 714  
DB 178 AAAAAAATTTACGAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTTACGAAAAA 119  
QY 715 TCGATGTTTAATGAGAAACAGCAGTCAATATAGCAATTAATTCAAAAATTAAGTTGAGC 774  
DB 118 AAATTATTAATAAAAAAATTTACGAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTT 59  
QY 775 ATAGTAATCAGATGATCTAGTAA 798  
DB 58 ATAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTTATTAATAAAAAAATTT 35

RESULT 4  
ABQ15589  
ID ABQ15589 standard; DNA: 778 BP.  
XX  
AC ABQ15589;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 2180.  
XX  
KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX  
OS WO200218632-A2.  
XX  
PN 07-MAR-2002.  
XX  
PD 01-SEP-2001; 2001WO-EP10074.  
XX  
PF 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.









KW Cowpox virus; bsr: viral vector; expression; apoptosis; resistance;  
 KW crm1; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;  
 KW autoimmune disease; graft rejection reaction; inflammation;  
 KW inflammatory disease; ss.

OS Synthetic.

OS Homo sapiens.

PN W09913073-A2.

PD 18-MAR-1999.

PF 07-SEP-1998; 98WO-JP04010.

PR 08-SEP-1997; 97JP-0259235.

PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.

PI Hamada H;

DR WPI; 1999-243728/20.

PT New apoptosis-resistant virus-sensitive cell

PS Example 3; Page 46-49; 51pp; English.

CC The present invention describes an apoptosis-resistant virus-sensitive  
 CC cell line into which an apoptosis resistance gene has been introduced.  
 CC The recombinant viruses generated are capable of expressing apoptosis-  
 CC associated genes. These can then be used in a variety of diseases for  
 CC which the induction of apoptosis is therapeutic. The recombinant viruses  
 CC are useful as vectors for gene therapy which can be applied to cancer  
 CC therapy for destroying cancer cells selectively, the treatment of  
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
 CC encountered the problem where if an adenovirus vector capable of  
 CC expressing an apoptosis-associated gene is introduced into animal cells,  
 CC the cells producing the virus will be destroyed because the period of  
 CC time required to induce cell death by apoptosis is shorter than that  
 CC required to replicate and produce the virus, resulting in failure to  
 CC obtain a recombinant virus having the integrated apoptosis-associated  
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
 CC apoptosis resistant gene introduced) is established and overcomes the  
 CC problem. The present sequence represents the base sequence of the  
 CC plasmid pR-Bcl 2-1-hCD 25, which contains the human Bcl-2 gene, and  
 CC is used in an example from the present invention.

CC SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 other:

Query Match 5.6%; Score 93.6; DB 20; Length 7996;

Best Local Similarity 45.9%; Pred. No. 9.4e-08;

Matches 321; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

OY 7 ACACAGCGCAATTTGAAAAATAGATTAATTTTGCAGCTATTAAAGCCGCTTAAC 66  
 DB 5139 AA 5198  
 OY 67 AAATGAGTGAAGAGAAAGAAATTAATACATTTTGTAGTAAAGAGAAAG 126  
 DB 5199 AA 5258  
 OY 127 AAAAAATTAAGACGCTCGGCTTACAGCTGGAAGAAACAGAAATAATTAAGAGAA 186  
 DB 5259 AA 5318  
 OY 187 GACTGTGATTTTATGGAATCGTGAGGAAAGAAATTTTAAATTTTCATTTTCGAGG 246  
 DB 5319 AA 5378  
 OY 247 ATTAATTTGTGTAGTGAAGAAATCTAGATTAATAATTCAGATCAAAATGCTTG 306  
 DB 5379 AA 5438

OY 307 AATTTGACATTAATGAAATACGTAGTAAATCAATTAATGAGGTTTGTCTATTATTG 366  
 DB 5439 AA 5498  
 OY 367 CGAAGATTGAAATCTGAGTGAAGAAATTAAGTTCGAGAGCAAAAAACCTTGCCGT 426  
 DB 5499 AA 5558  
 OY 427 TTTTCAATGACTTTGGAATAAATTCATGTGTGAGCGGTAGCAACTTGAAATTTT 486  
 DB 5559 AA 5618  
 OY 487 TACATTGGAATTTGAAAAATTAAGCAAGAACTCAATGAAAAATTTAATTATA 546  
 DB 5619 AA 5678  
 OY 547 AAAAAAGAGATCGATATGATTTTAAAGACGAAACTGCAATTGAAATGAAAAAG 606  
 DB 5679 AA 5738  
 OY 607 ATTGGAATAAATCTATGCTGAGAGTGAATTAAAGCAAAAAATTTGGCACTCAACCG 666  
 DB 5739 AA 5798  
 OY 667 GTGTGTTTTCAGAAATGACGATGAAGAAATGATGAAAAA 706  
 DB 5799 GAGATTTTCAGAGAGCTAAGGAAGCTAAATGAGAAAAA 5838

RESULT 9

ABL33958/c

ID ABL33958 standard; DNA; 14006 BP.

XX ABL33958;

DE 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1931.

KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antileukemia; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antithyroid; antidiabetic; antiparasitic;  
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

OS Homo sapiens.

PN W0200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation

PS Claim 1; SEQ ID NO 1931; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated



Db 2932 AA 2873  
Qy 487 TACATGGAAATTTGAAAAAAAAAAGCAAAAGAACTCAATGAAAAAAAAATTTATTATA 546  
Db 2872 AA 2813  
Qy 547 AAAAAAGAGATCGGATTTGATTTTAAAGCAGAAACCTGATTTGATTAATAAAG 606  
Db 2812 AA 2753  
Qy 607 ATTGGAATAATCTATGCTGAGATGATGAATAAAGCAAAATTTGGAACTGACCCG 666  
Db 2752 AACAAAAAACCAA 2693  
Qy 667 GTGTTGTTTGAATGACGATGAAGAAATGATGAATAATTCACCTCGATGTTATG 726  
Db 2692 AAAAAAAAAAAAAAAAAACAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2633  
Qy 727 AAGAAACGACGATCATATGAAATTTATTCAAAAATTAAGTTGACATGTAATCAG 786  
Db 2632 AACAAACAAAAAAAAA 2573  
Qy 787 A 787  
Db 2572 A 2572

## RESULT 11

ABQ29508/c  
ID ABQ29508 standard; DNA; 975 BP.

AC ABQ29508;

DT 12-JUL-2002 (first entry)

XX oligonucleotide for detecting cytosine methylation SEQ ID NO 16099.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetlig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

XX for diagnosis and prognosis, comprises selective hybridization of

XX amplicons from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CPG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

XX cytosine (C) but not methylated C, to uracil, then part of the genomic

XX DNA that contains the target C is amplified to form a labeled amplicon.

XX The amplicon is hybridised to two classes, each with at least one

XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

XX and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 975 BP; 112 A; 32 C; 51 G; 780 T; 0 other;

XX Query Match 5.3%; Score 88.2; DB 24; Length 975;

XX Best Local Similarity 45.9%; Pred. No. 7.1e-07;

XX Matches 379; Conservative 0; Mismatches 438; Indels 8; Gaps 2;

Qy 22 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 81

Db 850 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 791

Qy 82 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 141

Db 790 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 731

Qy 142 ACCTGCGCTTACGATGCAAAACGAGAAATTAATAAAGAAAGAGATGATTTTAA 201

Db 730 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 671

Qy 202 TGAATTCGATGAGGAGAAAGAAATTTTAAATTTTCGAGGATTAATTTGTTAA 261

Db 670 ATAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 611

Qy 262 GTGATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 321

Db 610 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 551

Qy 322 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 381

Db 550 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 491

Qy 382 TGAATTCGATGAGGAGAAAGAAATTTTAAATTTTCGAGGATTAATTTGTTAA 441

Db 490 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 431

Qy 442 TTGAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 501

Db 430 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 374

Qy 502 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 561

Db 373 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 314

Qy 562 ATATGATTTTAAAGCAAGAACTGACATTTGAAATGAAAAAAGATTTGAAAAATCT 621

Db 313 ATAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 254

Qy 622 ATGCTGAGAGGATTTAAAGCAAAATTTGGGAATCTCAACCGGTTGTTTAAAGAA 681

Db 253 AAAAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 194

Qy 682 TGACGATGAAGAAATGATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 741

Db 193 TAA-----AACCAACGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 139

Qy 742 AATATGAGAAATTTTCAAAATTAAGTTGAGATTAATTCAGATGATCTACTAAGCG 801

Db 138 ATAAATGATTAATTTTCGACGATTTTAAAGCCGACTTAAACAAATGAGTAAGAG 79

Qy 802 GACTATGAGTGTGGAACGATGATTTTGAATTAACCAAGAA 846

	DB	78	AAAAATTAATTAAAAAATAAAATAAAATAAAAAAAAAA	34
	RESULT 12			
	ID	ABQ29509	standard; DNA; 975 BP.	
	XX			
	AC	ABQ29509;		
	DT	12-JUL-2002	(first entry)	
	XX			
	DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 16100.		
	XX			
	KW	Human: cytosine methylation; 5'-Cpg-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; respiratory system; single nucleotide polymorphism. SNP; cell differentiation; ds.		
	KM			
	RN	Homo sapiens.		
	OS			
	XX			
	PJ	WO200218632-A2.		
	PD			
	XX			
	PF	01-SEP-2001; 2001MO-EPI0074.		
	PR	01-SEP-2000; 2000DE-1043826.		
	PR	05-SEP-2000; 2000DE-1044543.		
	PA	(EPIC-) EPIDEMICS AG.		
	PL	Olek A, Piepenbrock C, Berlin K, Gueutig D;		
	DR	WI; 2002-371829/40.		
	PT	Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA -		
	PS	Claim 12; 56pp + Sequence Listing; 56pp; German.		
	XX	This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (1) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNPs); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABO131410-BOQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.		
	XX	Sequence 975 BP; 780 A; 51 C; 32 G; 112 T; 0 other;		
	Query Match	5.3%; Score 88.2; DB 24; Length 975;		
	Best Local Similarity	45.9%; Pred. No. 7.1e-07;		
	Matches	379; Conservative 0; Mismatches 438; Indels 8; Gaps. 2		
	Db	22 AAAAATAGATAAATTTTCAGCATTAATGAACCGCACTTAAACAATGATGAAGAAG 81   126 AAAAATATTAATAATTAACCAACACATATTAATAATTAATAAACCAACGAATTAANA 185  82 AAAGAAAATAATATCATATTTTGAGTTGTGTAAGAGAGAAAGAAAATAATTAAGAAG 141		

Dd		186	AATTAATAAAAAAAAAAAGCATTAATAAAATTAAAAATTAACAACGATTAATAATA	245
Oy		142	ACCTGGCCTTACAGTCGAAAAACCAGAAATTAATAAAAAGAAGACTGTGATTTTAA	201
Dd		246	AAAAATTAANAACAAGCAATTAATAAATTAATAAATTAATAAANAACAATTAATAA	309
Oy		202	TGCAAATCGTAGAGAAAGAAATTTTAATTTTCATTTCCAGGGATTAAATTTGTGA	261
Dd		306	ATATAAAAAAAAAACCAATTAATAAATTAATAAACAACGATTAATAAATTAACGTAA	365
Oy		262	GTTGATCAAAAATCTGATTAATAAATAGCATCAAAAATGTGTGAATTGACATTTATG	321
Dd		366	AAACAAACGATTAATAAATTAATAAATTAATAAACAATTAATAAATTAATAA	423
Oy		322	AAATACGTACTATATCAATTAATGGGGTTTGTCTATTTTATTTTCGAAAGTTGAATC	381
Dd		426	AAACGATTAATAAATTAATAAATTAATAAACAACGATTAATAAATTAATAA	485
Oy		382	TGAGTCAGAGAAATATGTTGCCAGACGCAAAAACCCCTGCCGTTTTTTCCAATGCT	441
Dd		486	GAATTAATAAATTAATAAATTAATAAACAACGATTAATAAATTAATAAACAACGAT	545
Oy		442	TTGGAAAAAATTCATTTGTAGCGGTGCGAACCTTGAAATTTTTCATTTGAAATTTG	501
Dd		546	AAAAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAACAACGAA--TA	607
Oy		502	AAAAAATTAACGAAACCAATCAATGAAAAAATTTATTTAATAAAGAAGATCGG	561
Dd		603	AAAAATTAATAAATTAATAAACAACGATTAATAAATTAATAAATTAATAAACAACGATTAATAA	666
Oy		562	ATATGATTTTAAACGAGAAACCTGACATGTAATGAATAAATAAATGTTGAAAAAATCT	621
Dd		663	ATTAATAAATTAATAAACAACGATTAATAAATTAATAAATTAATAAACAACGATTAATAA	722
Oy		622	ATGCTGAGATGATTAATAAAGCAAAAAATTTGGGAATCTCAACCCGGTGTGTTAAGAA	681
Dd		723	AAAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAACAACGATTAATAA	783
Oy		742	AATATAGCAAATTAATTCAAAAATTAAGTTGAGCATATGTAATCAGATGATCTAGTAACG	801
Dd		838	ATATAAATAAATTAATAAATTAATAAATTAATAAACAACGATTAATAAATTAATAA	897
Oy		802	GACTATTTAGCTGGAAGCTCGAAATAGTTTGATTAATAAACAAGAA	846
Dd		898	AAAAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAA	942
<b>RESULT 13</b>				
AS4S4668/c	AD	AS4S4668 standard; DNA; 7442 bp.		
XX	AC	AS4S4668;		
XX	DT	18-DEC-2001 (first entry)		
XX	DE	Tumour suppressor gene derived chemically modified sequence #409.		
XX	KW	Human; tumour suppressor gene; oncogene; antitumour; cytostatic;		
XX	RW	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;		
XX	XX	Cytosine methylation; ds.		
OS		Homo sapiens.		
PN		MO200168912-A2.		
PD		20-SEP-2001.		
PF		15-MAR-2001, 2001WO-EPO2955.		





CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.

XX Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 other;

Query Match 5.0%; Score 83.8; DB 22; Length 9539;  
Best Local Similarity 43.0%; Pred. No. 5.3e-06;  
Matches 463; Conservative 0; Mismatches 612; Indels 2; Gaps 1;

```
QY 4 TCTACACAGCCAAATTTGAAAAATAGATAAATTTTCGACGTATTAAAGCCGACTTAA 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1116 TCTTCTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1057
QY 64 AACAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1056 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997
QY 124 AAGAAAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 996 TAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTT 937
QY 184 AGAGCTGTGATTTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 936 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 877
QY 244 GGGATTAATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 876 ACTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 817
QY 304 TTGAATTTGACATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 816 TACCCATTTTCTACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
QY 364 TTGCGAATTTGAAATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 756 TCCCACTACTCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 697
QY 424 CGTTTTTCAATGACTTTGGAAAAAATTCATTTGAGCGGTAGCGAACTTTGAAAT 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 696 AACGAAATCGTACACACTACACCTAACACAAACAAACAAACAAACAAACAAACAA 637
QY 484 TTTTACATTTGAAATTTGAAATTAAGCAAAAGAACTCAATGAAATTAATTAAT 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 636 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 577
QY 544 ATAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 603
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
QY 604 AAGATTTGAAAAATCTAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
QY 664 CCGGTGTGTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 723
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 397
QY 724 ATGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 783
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 337
QY 784 CAGATGATCTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 843
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 AATAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 277
QY 844 GAAGTCTTTCGTTTGTATTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 903
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 217
QY 904 ATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 ATACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 157
```

```
QY 964 TGAGTTTGTGTTTGTGATTAAGATTTTGTGAGTGAAGAGAGAGAGAGAGAGAGAG 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 99
QY 1024 CTCACACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 ATAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 42
```

Search completed: July 17, 2003, 05:07:52  
Job time : 392.725 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 02:22:20 : Search time 680.973 Seconds  
(without alignments)  
5635.000 Million cell updates/sec

Title: US-08-876-132-2  
Perfect score: 1847  
Sequence: 1 AGATCTCAACGACGTTTAAAA.....AACAAGCAAAAGAGATCT 1847

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 1448851 seqs, 1038787357 residues

Total number of hits satisfying chosen parameters: 2897702

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq3.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81.6	4.4	5689	US-10-239-676-90	Sequence 90, Appl
2	77.6	4.2	11812	US-10-239-676-210	Sequence 210, Appl
3	74.6	4.0	9539	US-10-239-676-51	Sequence 51, Appl
4	74.4	4.0	755	US-10-027-632-128109	Sequence 128109,
5	74.4	4.0	9515	US-10-239-676-160	Sequence 160, Appl
6	73.2	4.0	8085	US-10-172-086-40	Sequence 40, Appl
7	72.8	3.9	3252	US-10-027-632-113786	Sequence 113786,
8	72.8	3.9	3252	US-10-027-632-113787	Sequence 113787,
9	72.6	3.9	3252	US-10-027-632-113788	Sequence 113788,
10	72.6	3.9	9539	US-10-239-676-95	Sequence 95, Appl
11	72.6	3.9	15732	US-10-239-676-28	Sequence 28, Appl
12	72.6	3.9	17848	US-10-239-676-96	Sequence 96, Appl
13	72.6	3.9	15732	US-10-239-676-143	Sequence 143, Appl
14	69	3.7	8801	US-10-239-676-18	Sequence 18, Appl
15	67.8	3.7	5979	US-09-790-988-1	Sequence 1, Appl
16	67	3.6	640681	US-09-790-988-1	Sequence 1, Appl

17	66.4	3.6	335913	US-09-754-853A-2	Sequence 2, Appl
18	66.4	3.6	335913	US-09-754-853A-3	Sequence 3, Appl
19	66.2	3.6	7195	US-10-239-676-30	Sequence 30, Appl
20	66	3.6	53332	US-09-801-861-3	Sequence 3, Appl
21	66	3.6	53332	US-10-224-562-3	Sequence 2, Appl
22	66	3.6	335913	US-09-754-853A-2	Sequence 3, Appl
23	66	3.6	335913	US-09-754-853A-3	Sequence 3, Appl
24	65.6	3.6	960	US-10-198-846-6381	Sequence 6381, Ap
25	65.6	3.6	271990	US-10-195-144-87	Sequence 87, Appl
26	64.8	3.5	11047	US-10-239-676-188	Sequence 188, App
27	64.8	3.5	11836	US-10-239-676-102	Sequence 102, App
28	64.6	3.5	12968	US-10-239-676-202	Sequence 202, App
29	64.4	3.5	516	US-09-960-352-5785	Sequence 5785, Ap
30	64.4	3.5	11836	US-10-239-676-101	Sequence 101, App
31	64.2	3.5	17421	US-10-239-676-54	Sequence 54, Appl
32	64	3.5	1431	US-09-774-414-2	Sequence 2, Appl
33	63.8	3.5	2086	US-10-027-632-250719	Sequence 250719,
34	63.8	3.5	2086	US-10-027-632-250720	Sequence 250720,
35	63.8	3.5	5979	US-10-239-676-18	Sequence 18, Appl
36	63.6	3.4	431	US-09-960-352-5558	Sequence 5558, Ap
37	63.4	3.4	628	US-10-027-632-113152	Sequence 113152,
38	63.4	3.4	6665	US-10-239-676-4	Sequence 4, Appl
39	63.4	3.4	14147	US-10-172-086-51	Sequence 51, Appl
40	63.2	3.4	1223197	US-10-027-632-179264	Sequence 179264,
41	63	3.4	2053	US-10-017-128-3	Sequence 3, Appl
42	63	3.4	159095	US-10-017-128-3	Sequence 3, Appl
43	62.6	3.4	1678	US-10-155-533-6	Sequence 6, Appl
44	62.6	3.4	1910	US-10-155-533-2	Sequence 2, Appl
45	62.6	3.4	5314	US-10-155-533-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-10-239-676-90/c  
; Sequence 90, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239, 676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 90  
; LENGTH: 5689  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; NAME/KEY: unsure  
; LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)  
; US-10-239-676-90

Query Match 4.4%; Score 81.6; DB 14; Length 5689;  
Best Local Similarity 45.2%; Pred. No. 0.00062;  
Matches 400; Conservative 0; Mismatches 477; Indels 7; Gaps 3;  
QY 907 TTTTATACGACGACCAAAATTTTAAACAAAGAAATTTTCGACGACAAAT 966

D	b		2074	TTTTTCCTCGGAAAAAAAAAAAAAAAAATATATAAAAAAAAAACAAATATAAAAAAAAAAATATATTT	2015
O	y		967	ATTTTGGAATTTTTTAAAGCGATCTTGTCACGCGCACTTTTGCCATATTTTAAACCTGA	1026
D	b		2014	ATTTTTATATATTTTTTTTTTTTTTATTTTAAAAAATATTTAAAAAANATATAAAAAATTT	1955
O	y		1027	CTATCTTATAGTTAATAGATATATCGGTAGATTATAAGGTATGTTAAAAACGAGTA	1086
D	b		1954	TAAATDANNTTAAAAAANNTANNATAAAAAANATATATATATAATAAAAAATTTATTT	1899
O	y		1087	AAACATATACCTTATATTTATTTCTGATTTATATTTCACAGTGTATTTTAAATATATTA	1146
D	b		1894	AATATAAAAAATATAATATAAAAAAATATAATTTATTAATTTTATTTATATATAT	1835
O	y		1147	AGAGATATATCTATTAGCTTAATATACTAAAAAAGAGTAAATATATGATTTGTA	1206
D	b		1834	TAAATATAAAAAATTTANTAAAAAATTTATAAAAAATATAAAAAAATATATTAATAA	1757
O	y		1207	----TTTAAAAAGCATTAGAAAAATGAATAGAACATTTATAAAAAAGCGTATATCAA	1263
D	b		1774	ATATATATAAAAAAATATAATATATAATAATAATAAAAAAATATAAAAAATATTA	1715
O	y		1263	ATCTTTCTTACAAATCTTGCACTTACTTGTATATAGATTAACCATTAATGGTGATGATG	1322
D	b		1714	AAAAATAAAAAAGAAAAAATATAATACAAATATAAAAAATACCAATATTTTATTCATTA	1655
O	y		1333	TGGCGATATTAACATCBAATTTATTCATTGATGAAGTTTGTCTTCTTGATGTTGA	1382
D	b		1654	ATTAATAACAAAAAACAATAAATATATATCTATCAAAAAATA--AATAAAAATTAATTA	1597
O	y		1383	GCACAATTTGGGCTGGCGCTTCATATAAATAATATACAGACAGAAATGATATTTATCTC	1442
D	b		1596	AAATATATACAAATTTTATATAAAAAAATATAAAAAATTTTATATACCTTAAAAATATA	1537
O	y		1443	AGCTMAAATGAAATGATTTTAAAAAATACCTGTTCTTTATTTTCATTTGATCAATGGGA	1502
D	b		1536	ATATATACCTATATAAAAAACACATTCGATATCAAAATCGTATACAAATTTATCTCAGCAT	1477
O	y		1503	AGATTTTATATTCGATGATGAGTACATATCATCACAAGAAATGATTAAGATTTAGATTA	1562
D	b		1476	AATATAAATAAAAAAATATATATTCACAATAAATATATCAAAATATAAAAAATTTATTTT	1417
O	y		1563	TTACACATATGAGAAATTAACCATATGATATATAAATAAGAAATATATGATTTGCA	1622
D	b		1416	ATATATTAATATTTATTAATAATACGAAATTTATTTACATAAAAAAATATATATTTATA	1357
O	y		1623	ATAACAAAGATATATAAAGAAAGTCAGCTTTAAAAAATATCGATTGATGATGTTT	1682
D	b		1356	TAAATATTAACAAATATATTTACTACATTAATATTTCAATATATAATTTAATATATAATTA	1297
O	y		1663	AAAGTAGATGA--AAATTTATTTGATGAAGAGATTTGCACTMAAAGATATTGAAATATC	1741
D	b		1296	AAAAATAAAAAATAACATATATTTCACTTTACAAATTCATATTTAAATATATAAAAAATATAT	1237
O	y		1742	GTAAGAAATCCATTTGATGATCCAGATGATGGCATTAAGAAAT	1785
D	b		1236	ACTTATATATAATAATTAATTAACCTTTTATTTAAAAACTAATAT	1193

```

RESULT 2
US-10-239-676-210/c
; Sequence 210, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: Pct/EP01/03968

```

```

? DE 10019058.8
? DE 10019173.8
? DE 10032529.7
? DE 10043826.1
? PRIOR FILING DATE: 2001-04-06
? 2000-04-06
? 2000-04-07
? 2000-06-30
? 2000-09-01
? NUMBER OF SEQ.ID NOS: 228
? SEQ ID NO 210
? LENGTH: 11812
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-210

```

Query Match	4.28;	Score 77.6;	DB 14;	Length 11812;
Best Local Similarity	45.68;	Pred. No. 0.003;		
Matches 311;	Conservative	0;	Mismatches 369;	Indels 2;
				Gaps 1.

Oy	1056	TTGACATTTAAAGATGTTGTTAAAGACGATTAAGACATTAAGCTTTATTTTAAATTCGAA	1115
Db	2531	TTATATTTAAATACACAAAAAACAACAAAAAAATCTCTATTCATTTAAAAAA	2472
Oy	1116	TTATATTTGACAGTGATTATTTAAATATTTAAGAGATATCTATTTAGCTTTAAATATTAAC	1175
Db	2471	TAAATATATTTTCAAAAAATTAATTAACGACGAACTAACAAACTATATACCTTAANAATCGAA	2412
Oy	1176	TAAAAAAAAGCGTAAATATATGATTTGTGTTTAAAAAACGATAGAAAAATGAAATGAA	1235
Db	2411	TAAACACCTATTTTATTTATTTAAATTTTAAACGTAAAAATATTTATACAAAAAACATTTA	2352
Oy	1236	ACATTATATAAAAGACGGTGATATCAAAATCTTTCTTACATATCTTGCACTTACTTGATAT	1295
Db	2351	CAACCATTTAAACAATTAATTAATTAATTAACCTTTATACCCCAATTAATTAATTAATTTCTCT	2292
Oy	1296	AGATTAAGCATTAATGGTGTGATGATGGCGATATTTAAACGCAAAATTTATGCATTTGA	1355
Db	2291	TAAAAAAACATTTCTATTTCTTTATTAATTAACCTATATTTACAAAACCTTTACTATTTATAT	2232
Oy	1356	TGAAAGTTTGTGATCTTCTGTGATGTTGAGCACAAATTTGGCTGGGCTTTCAATTAATATAT	1415
Db	2231	TTCTATTTCAAAATTTTAAAAAATTAATAAATTTAATCTCGTTTTCATATAA - AT	2174
Oy	1416	ACACAGACGAATGAATTTATATATATGATACGCTAAACCTGAANAATGATTTAAAAAATCTC	1475
Db	2173	ACTATATATAATTTTCAATTAATTTACTTTAAATACAAAATCTAAAAATTTTACTTTCTA	2114
Oy	1476	GTTCCTTTATTCATTCGATTCATTTGGGAAGATTTAAATTACGATGAGTGAATACAATPACA	1535
Db	2113	CTTATTAACAAATTTAAAAAANCATATATTAATTTAAAAACAAAACATATACACTATAA	2054
Oy	1536	TCAAGCAATGATTTAAGAGTTAGTAAATTTACACATATGGAGAAATTAACATATGATATAT	1595
Db	2053	ATTAATAAAACCAAACTATTAATCTATTTATTAATAAAACCTAACAAATTTAAAAATTAATCT	1994
Oy	1596	AAATTAAGAGATATTAATGATTTGCAATATACAAAGATATATAAGAAAGACAGCTCTTA	1655
Db	1993	AATTAATAAAATTAAGAGTTAGCTTATTAATAAAATTTTATTTTAAATTAATATTTT	1934
Oy	1656	AAAAATATCTGCATTTGATCGATGTTTTTAAAGCTAGATGAAAAATTTATTTGATGAAGAGAT	1715
Db	1933	CTATATAAACAAAAATTAATAAAATTTTAAATAAATTTTATCAATATATATAC	1874
Oy	1716	TTGCACTAAAGATTTGAAA	1737
Db	1873	GTTCATTCATCAATACACA	1852

RESULT 3  
US-10-239-676-51/c

```
; Sequence 51, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: FIEBENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 51
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-51
```

```
Query Match 4.0%; Score 74.6; DB 14; Length 9539;
Best Local Similarity 45.6%; Pred. No. 0.0087;
Matches 381; Conservative 0; Mismatches 444; Indels 10; Gaps 3;
```

```
QY 1012 AATTTTAAACCTGATCTTTATAGTATATCGGTGATTAAGTAT 1071
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 6233 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6174
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1072 GTTAAAAAGAGTAAACATTAATATATATATATATATATAT 1131
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 6173 AACCATTAATTTATTAATAAATAAATAAATAAATAAATAA 6114
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1132 TATTTAATATATTAAGATATATATATATATATATATAT 1191
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 6113 AATTCAAAA---ACACATTTTAACTCAAAAATATATTA 6058
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1192 TATATGATGTGTATTTAAAAAGCATTAAGAAATGACATT 1251
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 6057 AAAATCGAAACGAAATTTAAAAATCAAAAATTAATAAT 5998
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1252 GGTATATCAAAATCTTTCAATCTGATTTGATATATAGAT 1311
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5997 TCTAAAAATCAATAAAAATCTAAAAATTTCTAAAAATTA 5938
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1312 GGTATGAAATGTGGCATTTATTAACGAAATTTTCCATGAT 1371
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5937 TAAAAATCAAAATCTATTAACAACAAATATATATAAAA 5878
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1372 CTTCATGTTGACACATTTCCGCTGGCTTCAATTAATAAT 1431
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5877 CCAAAATATAAATTAATAAATAAATAAATAAATAAATA 5818
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1432 TATTTATCATGCTGATTAAGTATTAATAAATATCCGCT 1491
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5817 AATATATTTTAAATCAATTAATTAATAAATAAATAAATA 5758
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1492 ATCAATTTGGAGAAATTTATTAAGATGATGATCAATCA 1551
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5757 ACAAAATTTTAAAAATTAATTAATTTAA--AATAACAA 5700
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1552 GGATTAAGATTAATTAACATATGAGAAATTAACATTA 1611
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5699 CAAAAATTAATAAATAAATAAATAAATAAATAAATA 5640
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 1612 TTGATTTCCAAATACAAAGATATAAGAAAGTCAGCTTA 1671
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5639 AATTAATTTTAAAAACAAAAATTAATAAATAAATAAATA 5580
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1672 TCGATGTTTTTAAAGTAGATGAAAAATTTATGATGAG 1731
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5579 AACAAATTTTAAACCTCAAAAAATAAACAACAAACGTT 5524
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1732 TGAATATATCGTAAATATCCATGATGATCCAGATG 1791
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5523 AAAAAATTTTAAACAAAAATAAATAAATAAATAAATA 5464
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1792 TCGCAGAGAAATGCGTATGCTTCCGATTAATAAACA 1846
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 5463 TTAATTAATTAATAAATCAATATATAACAAAAAAAT 5409
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 4
US-10-027-632-128109/c
; Sequence 128109, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128109
; LENGTH: 755
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-128109
```

```
Query Match 4.0%; Score 74.4; DB 14; Length 755;
Best Local Similarity 51.3%; Pred. No. 0.006;
Matches 199; Conservative 1; Mismatches 182; Indels 6; Gaps 1;
```

```
QY 967 ATTTTGAATTTTAAAGGATCTTCTACCGCATTTTGCAT 1026
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 752 AATTTTAAATATATATATATATATATATATATATATAT 693
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1027 CTATCTTTAATTAATATATATATATATATATATATAT 1086
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 692 TATTAATATATATATATATATATATATATATATATAT 633
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1087 AAACATTAATATATATATATATATATATATATATAT 1146
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 632 AATATATATATATATATATATATATATATATATATAT 573
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1147 AGATATATATATATATATATATATATATATATATATAT 1206
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 572 TATTAATATATATATATATATATATATATATATATAT 513
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1207 TTTAAAAAGCATTAATAAATAAATAAATAAATAAATA 1260
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
DB 512 TATTAATATATATATATATATATATATATATATATAT 453
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
```



[illegible]

```

RESULT 8
US-10-027-632-113787
; Sequence 113787, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113787
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-113787

Query Match      3.9%; Score 72.8; DB 14; Length 3252;
Best Local Similarity  44.9%; Pred. No. 0.014;

```

[illegible][illegible]







```

;      2000-04-07
;
;      2000-06-30
;      2000-09-01
;      NUMBER OF SEQ ID NOS: 228
;      SEQ ID NO 96
;      LENGTH: 15732
;      TYPE: DNA
;      ORGANISM: Artificial Sequence
;      FEATURE:
;      OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-96

```

Query Match	3.9%	Score 72;	DB 14;	Length 15732;
Best Local Similarity	44.9%;	Pred. No. 0.024;		
Matches 311; Conservative	0;	Mismatches 380;	Indels 1;	Gaps 1

QY	1012	ATATTTAAACCTGACATCTCTTTAAAGTTAATGATATATTCCTGATACATATATAAGAT	1071
Db	5130	ATATATATATAATGAATAATGATATAAGAAATTTATACGTTAAATTTATATATTTTGG	5189
QY	1072	GTTAAAAACGAGTAAAAACAATPACTATATATTTTAAATCTCGAATTAATATTTGACAGTGA	1131
Db	5190	AGAAATTAATAGAAATATATATATATTTTATTTAATGTAAAGAAATATGCTGATGAG	5249
QY	1132	TTATTTAATATATTAAGAGATATATCTATTTAGCTTAAATATATACTAAAAAAGAGCTAA	1191
Db	5350	GTAATTTATATACATTTTAAATAATTTATATTTATATAAGAAATTTTAAATTAATATTTGAT	5309
QY	1192	TATATGATGTGTATTTTAAAAAAGCATATAGAAAATGAAATAGAAACATTTATAAAAAAGAC	1251
Db	5310	TTTATATATATAGGAATTTTAAAAAAGTAGAATTAATTAATTTAAAGCTATTTAAAAAGGA	5366
QY	1252	GGTGATATCAAACTCTTCTTACAATPACTGTCATTACTTGATPATAGATTAAGCATTAAT	1311
Db	5370	GGAATAAATAAATAATTTAGAGTAAATTTGAGATAGAAAAATA-ATAAAGAAATTAATA	5428
QY	1312	GGTATGATGTGCGCATATATATAACCAATTTATTCATGATGAAGAAGTTTGATCTT	1371
Db	5429	AAAAAAGAGTGGTTTTTTTGAATAATGTGAATATATTTGATAGAAATTTTATTTGATTTGA	5488
QY	1372	CTTGATGTGGACACAATTTGCGCTGGGCTTTCATATAAATPATACAGACGAAATGAA	1431
Db	5489	TATGGAAGAAAAAAGAGAGATTTAAATATAAATAATTTGAAATTTAAAAAGGAATATTTAT	5548
QY	1432	TATTTATCATGAGTAAACTGAAATATTTTAAAAAATCTCGTCTTATATTCATTCG	1491
Db	5549	AATAGATTTTATAGAAAATTAAGAGAAATTTTGTGAATTAATTTATATTAATTAAGTGGATA	5608
QY	1492	ATCAATTTGGAGAAATTTAATTACGATGAGATGAGTACATATACATCAAGAAATGATTTAA	1551
Db	5609	ATTTTAAATGAAATGCAATTAATGCAATTAATTTATTTGAAATTAATTTATTTGTTAAATTTTGAA	5668
QY	1552	GGATTAGATATTTACACATATGSGAAATATAACCATATGCAATATAATTAATTAAGAGATATA	1611
Db	5669	TTATATAAGAAATAGAAATTTGCAATAGTTATATGATTAATAGTTAGAGATTTGAATTTAGAG	5728
QY	1612	TTGATTTTGAATATAACAAAGATATATAAAGAAAGTCAAGCTTAAAAATATCTGCATTTGA	1671
Db	5729	TTAAAAATTTTATTATATAAGAAAGTTTAGATTAAGTGGTTTTATTTGGTAAATTTTATATA	5788
QY	1672	TCGATGTTTTAAAAAGTAGATAAAAATTTATT	1703
Db	5789	AATATATATAAGAAATTTAAGATTAATAAATTTT	5820

RESULT 14  
US-10-239-676-143/c  
: Sequence 143, Application US/10239676  
: Publication No. US20030082609A1  
: GENERAL INFORMATION:  
: APPLICANT: OLEK, Alexander  
: APPLICANT: PIERDENROCK, Christian  
: APPLICANT: BERLIN, Kurt

```

1      TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
2      FILE REFERENCE: 5013.1003
3      CURRENT APPLICATION NUMBER: US/10/239,676
4      CURRENT FILING DATE: 2002-09-24
5      PRIOR APPLICATION NUMBER: PCT/EP01/03968
6      DE 10019058.8
7      DE 10019173.8
8      DE 10032529.7
9      DE 10043826.1
10     PRIOR FILING DATE: 2001-04-06
11     2000-04-06
12     2000-04-07
13     2000-06-30
14     2000-09-01
15     NUMBER OF SEQ ID NOS: 228
16     SEQ ID NO 143
17     LENGTH: 8801
18     TYPE: DNA
19     ORGANISM: Artificial Sequence
20     FEATURE:
21     OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
22     US-10-239-676-143

```

Query Match	3.7%	Score 69	DB 14	Length 8801
Best Local Similarity	46.7%	Prod. No. 0.065		
Matches 356	Conservative	0	Mismatches 400	Indels 7
				Gaps 4
QY	TTATACCGGAGACAAACAAAAAATAAACAAGAAAAATTTTCGACGAAAAAAATTT	969		
Db	TTAAAAAATAAATAAACCACCAACCAAAACACACAAACACCTTTTTCATATACAAATATA	2246		
QY	970 TTGCAATTTTAAAAGCGCACTTTCGACGCACTTTGGCATTATTTAAACCTGACTA	10292		
Db	CTATATTTTAACTACATCTTTCTCTAAATATCATAAAAACATCTCTTAAATATACCA	21866		
QY	1030 TCTTATATAGTATATAGATATATTCGCTAGATATATAAGTATATGTTAAAAAGAGATAAA	10899		
Db	2185 TCTAACAAAAAATAATTTTAAAAAACAATATATCTATATAATCAATATTTCACACTA	21282		
QY	1090 CAATTAATTAATATATTAATTCGTGATATATTTGCAGATGATTAATTAATATTAAGA	11499		
Db	2125 AATAAATAAATAATATTTTAAATTAATATACAAAC--TAAATATTTAAAAATAACT	20688		
QY	1150 GATAATCTATTAAGCTTAATATATTAATTAACATAAAGAGATAATATATGATTTGTATTT	12099		
Db	2067 AATTTAATTAATTCATCAAT--TAAATATCAACATTAATTAATTAATTTCTCATCATTT	20099		
QY	1210 AAAAAAGCATAGAAAAATGAATGAGACATTATAAAAAAGACGGTGATTCMAATCTTTC	12693		
Db	2008 ATACTTCATCATCGAATATATATATAAATAAATAAATAAACAATCATATATATTTTC	19494		
QY	1270 TTACAAATCTTGCACTACTTTTGATATAGATTAAGCATTAATTTGGTGATATGTGGCAT	13292		
Db	1948 TACTTAAACAACATTAATTTTAAAAATTTACTTAAACACAAACCATCACTCATCT	18899		
QY	1330 ATTATAACTCAAAATTTATCATGATGATGAAGTTTGATCTCTTGATGTGGACAAAT	13899		
Db	1888 TTTATTAATATATTTTCTATATTTTCAACAATTCACAAATTTTAAACACACCATTAT	18292		
QY	1390 TTGGCTGGCTTTCAATTAATAATAACAGAGACGAATGAATATTTATCATCACTAAA	14499		
Db	1828 ATTACATTCACCAAAAAAATAAATACTACCAATAAACATATATATCCCTTAAATATAA	17695		
QY	1450 ACTGAAAATGATTTTAAAAAATACCTCGTCTTTATCTTCATGATCAATTTGGGAAGATTT	15092		
Db	1768 ATACCTCTTATTTCAAAAAAATAATTAATTAATTAATTAATTTCTTAACTTAATA	17099		
QY	1510 AATTACGAT--GAGATGAGTACATATACATCAGAAATGATTTAAGGATAGATA--ATA	15695		
Db	1708 AAATATATATATCAATACCTTCATTAATCTTATATATATCTCATATATATTTATATATA	16499		
QY	1566 CACATATGAGAAATTAACCATATGATATATAAATAGAGATATATGATTTGCAATA	16295		

DB 1648 CACAAATACCTTAACCTTCAAAATTAACCTATCTTAATAAATTAATCTCTTAAATATA 1589  
 QY 1626 ACAAAGATATATAAGAAAGTCAGCTCTTAAAAATATCTGCAT 1668  
 DB 1588 AAAAAACATATAACGATATCCCTTTTATTAATAATCATCTTTAT 1546

## RESULT 15

US-10-239-676-18/c  
 ; Sequence 18, Application US/10239676  
 ; Publication No. US20030082609A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OLEK, Alexander  
 ; APPLICANT: PIEPENBROCK, Christian  
 ; APPLICANT: BERLIN, Kurt  
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
 ; FILE REFERENCE: 5013.1003  
 ; CURRENT APPLICATION NUMBER: US/10/239,676  
 ; CURRENT FILING DATE: 2002-09-24  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
 ; DE 10019058.8  
 ; DE 10019173.8  
 ; DE 10032529.7  
 ; DE 10043826.1  
 ; PRIOR FILING DATE: 2001-04-06  
 ; 2000-04-06  
 ; 2000-04-07  
 ; 2000-06-30  
 ; 2000-09-01  
 ; NUMBER OF SEQ ID NOS: 228  
 ; SEQ ID NO 18  
 ; LENGTH: 5979  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 US-10-239-676-18

Query Match 3.7%; Score 67.8; DB 14; Length 5979;  
 Best Local Similarity 45.4%; Pred. No. 0.094;  
 Matches 358; Conservative 0; Mismatches 427; Indels 4; Gaps 3;

QY 1043 ATAGATATATCCGTAGATTAATAAGTATGTTAAAAAGAGTAAAAACATATATAT 1102  
 DB 3900 ATATATATTTTTTTTTTCCAAATCCCTTAAATTCCTTCTATATTAATTTCTTT 3841  
 QY 1103 ATTTATCTGATATATATTTGACAGTGATTAATTAATATTAAGAGATATATCATTA 1162  
 DB 3840 AAAAAACTTAAAAAATTTCTACATTAATAATTTAAATCCACATATTTCTATATTTTA 3781  
 QY 1163 GCTTAATATATTAACATAAAAAAGGTAAATATATATATGATTTGTTAAAAAGCATTA 1222  
 DB 3780 ACTTCCAAAAATTTTAATTTTAATTTTATTTTACTTTTATATACATAATATA 3721  
 QY 1223 AAAATGAATAGACATTTATAAAAAGACGGTGATATCAATCTTTCTACATATCTTGC 1282  
 DB 3720 ATATTAATATTAATAATTAATAAATCTTAACCTTCACTAAAACTTTTATTAATAACAAT 3661  
 QY 1283 ATTACTTGATATGATTAAGCATTAATGATGATGATGCGATATTTATAAATCA 1342  
 DB 3660 ATATCTATATATATCAAAAACTTTATTTACTTTAAAGTACATCTTATCAAAAAATTA 3601  
 QY 1343 ATTATCCATGATGAAGTTTGTATCTTCTTGATGTTGAGC-ACAATTTGCGTGCGCT 1401  
 DB 3600 TATATATTTTACATACACTTTACAAATTTCAATTAATAAACTATTAACAAATTAATAAT 3541  
 QY 1402 TTCAATTAATATATACAGAGAGAAATGATATTTATCATCATGCTAATAACGAAATGAT 1461  
 DB 3540 TTATATATATTAATAATATCTCAAAATTAATAATATTTTCTATTTAAAAAATTAATA 3481  
 QY 1462 TTTAAATAATCTGCTGTTTATTTATCGATGATGGAAGATTTAATGATGAG 1521  
 DB 3480 CCACCAAGCTTTTCTTTTATATTTTTCATCA-ACACCAATCAACTTAATAAAT 3423

QY 1522 ATGACTACAAATACATCAAGAAATGATTTAAGGATTAGATTAATTTACATATGAGAAATA 1581  
 DB 3422 TCAATTAACCTTACCACTAAAAAATCAAAATTAACAAATCTACTATTAATAAATA 3363  
 QY 1582 ACCATATGATATTAATAATTAAGAGATATATGATTTGGAATTAACAAAGATATAAAG 1641  
 DB 3362 AATTAATTAATTAACAAAAATCTTATTAATTAATTAACAAACAAAAATTTCTTTAACA 3303  
 QY 1642 AAGTCA-GCTCTTAAAAATATCTGCATTTGATGATGTTTAAAGTATGATGAAAAATTT 1700  
 DB 3302 AATTCATATCAATTAATAATTAATAAATTTCTTACCAACAAATTAACATTAATTT 3243  
 QY 1701 ATTTGATGAAGAGATTTGCACTAAAGATATTTGAAATATGCTATGAAATCTATGAT 1760  
 DB 3242 AACTTCTTAATAAATCTTTTATTAACATTTTAACATTTTAACATTTAAAAAATAACACACT 3183  
 QY 1761 GATCCAGATGATGCAATGAAGAAATCACAATTCGACGAGAGAAATGCTATGCTTCCGC 1820  
 DB 3182 AACCCAAATTAATAATTTTATACCTTAATAAATTTTACAAATTTAATTAATTAATGATAAC 3123  
 QY 1821 ATTAATAA 1829  
 DB 3122 CTCTTAATA 3114

Search completed: July 17, 2003, 07:01:01  
 Job time : 691.973 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 02:19:35 ; Search time 89.0057 Seconds  
(without alignments)  
6363.996 Million cell updates/sec

Title: US-08-876-132-2  
Perfect score: 1847  
Sequence: 1 AGATCTCAACGAGTTTAA...AAAAACAAGCAAAAGATCT 1847

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: Issued\_Patents\_NA:\*  
2: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/6C\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.6	4.0	19124	2	US-08-487-826B-13
2	65.8	3.6	20674	4	US-09-641-638-651
3	65.2	3.5	3095	6	5231168-1
4	64	3.5	1431	4	US-09-316-083-2
5	63.2	3.4	7218	1	US-08-332-463-14
6	62.4	3.4	19124	2	US-08-487-826B-13
7	62.2	3.4	6243	2	US-09-056-075-1
8	61.6	3.3	665	2	US-08-883-795A-36
9	61.4	3.3	615	4	US-08-998-416-186
10	60.2	3.3	5852	1	US-07-867-106-2
11	59.4	3.2	636	4	US-08-998-416-1137
12	59	3.2	6124	4	US-08-213-419B-3
13	58	3.1	837	4	US-08-998-416-288
14	58	3.1	8920	2	US-08-446-855A-1
15	58	3.1	8920	4	US-09-150-741-1
16	57.8	3.1	4140	3	US-08-894-731-2
17	57.6	3.1	872	4	US-08-998-416-487
18	57.2	3.1	6124	4	US-08-213-419B-3
19	56.4	3.1	658	4	US-08-998-416-595
20	55.6	3.0	677	4	US-08-998-416-1092
21	55.6	3.0	821	4	US-08-998-416-541
22	54.8	3.0	2739	4	US-09-134-001C-156
23	54.4	2.9	665	2	US-08-883-795A-36
24	54.4	2.9	5852	1	US-07-867-106-2
25	54.4	2.9	20674	4	US-09-641-638-651
26	54	2.9	636	4	US-08-998-416-1137
27	54	2.9	8133	1	US-08-480-604A-5

28	54	2.9	8133	2	US-08-405-496A-5	Sequence 5, Appl
29	54	2.9	8133	4	US-08-915-136-5	Sequence 5, Appl
30	54	2.9	8133	4	US-08-957-310-5	Sequence 5, Appl
31	53.6	2.9	1689	1	US-07-991-867B-41	Sequence 41, Appl
32	53.6	2.9	1689	2	US-08-544-332-41	Sequence 41, Appl
33	53.6	2.9	1689	4	US-09-370-861A-41	Sequence 41, Appl
34	53.6	2.9	1947	4	US-07-991-867B-1	Sequence 74, Appl
35	53.6	2.9	8457	1	US-07-991-867B-1	Sequence 1, Appl
36	53.6	2.9	8457	2	US-08-544-332-1	Sequence 1, Appl
37	53.6	2.9	8457	4	US-09-370-861A-1	Sequence 1, Appl
38	53.6	2.9	290	3	US-08-617-860B-13	Sequence 13, Appl
39	52.8	2.9	1511	1	US-07-991-867B-8	Sequence 8, Appl
40	52.8	2.9	1511	1	US-08-107-755A-8	Sequence 8, Appl
41	52.8	2.9	1511	2	US-08-544-332-8	Sequence 8, Appl
42	52.8	2.9	1511	4	US-09-370-861A-8	Sequence 8, Appl
43	52	2.8	711	4	US-09-134-001C-1758	Sequence 1758, Ap
44	52	2.8	2435	4	US-09-306-593-1	Sequence 1, Appl
45	51.6	2.8	1956	4	US-08-559-896B-1	Sequence 1, Appl

# ALIGNMENTS

RESULT 1  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knodbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 4.0%; Score 74.6; DB 2; Length 19124;  
Best Local Similarity 46.8%; Pred. No. 1.9e-06;

Matches 310; Conservative 0; Mismatches 344; Indels 9; Gaps 2;

```

OY 1037 AAGTTAATAGATATATCCGTTAGATTATTAAGTATGTTAAAAACGATTAACATTAAC 1096
    || || || || || || || || || || || || || || || || || || || || ||
Db 15453 AAAAAAGAAAAATGAATATATAAAAAATTTATTAATAAAAAATAAAAA 15512
OY 1097 TTAATATATTTAATCTGAAATTAATTTGACAGTGAATTTAATATATTAAGATATAT 1156
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15513 AAGGAGAAAAATTTTAAAAAATATTAATAATTAATAATTAATAATTTTGATAGAA 15572
OY 1157 CTAATAGCTTAATTAATTAAGTAAAGGTAATATATATGATGTGATTTAAAAAG 1216
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15573 TAAAAAATGAAGAAAGATTATCAAAAAAATTAATAAATTTATATTAATAAATAAT 15622
OY 1217 CATTAGAAAAATGAATAGAACATTATAAAAAGCGTGATATCAAACTTTCTTCAAT 1276
    || || || || || || || || || || || || || || || || || || || || ||
Db 15633 GATTATATAAAAAATTAATAAACAAGAAAGAAAAAACAATTTAA-----AAAAAA 15686
OY 1277 ACTGCATTACTTGTATATAGATAAGCATTTAAATGGTGATGAATGTGGCGATATTATA 1336
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15687 AATATATATATCAATTAATAAACAAGAAAAAATTAATTAATAATTAATAATATATAT 15746
OY 1337 ACTCAATTTTATCAATTCATGTAAGTTTGTATCTTGTGATGTGACAAATTCGGCT 1396
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15747 CATTAATAATTAATAAATTAATAAATTTTAAAAAATTAATATATATTAATAAATAA 15806
OY 1397 GGGCTTTCATTAATAATATACAGACGAATGAATTTTATCATCAGCTAAAGTGAAA 1456
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15807 AATTTATTTAATTAATAAATAAATTAATAAATAAATTTAATTAATTAATAAATAA 15866
OY 1457 ATGATTTTAAAAAATATCTGTTTATTCATTCATCAATGGGAGAAATTAATTACG 1516
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15867 TAAATTAATAAATAAATTAATAAATAAATAAATAAATTTAATGACAAATAAATAA 15926
OY 1517 ATGAGATGAGTCAATACATACAGAATATGATTAAGATTAATTAATTAATTAATGAG 1576
    || || || || || || || || || || || || || || || || || || || || ||
Db 15927 ATAAAAAATTTAATTAATTAATAAATAAATAAATAAATTAATTAATTAATTAAT 15986
OY 1577 AATATACCATATGAATTAATAATAAGAAATATATGTTTGA--ATACAAAGAA 1633
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db 15987 ACATATATATATATATATACCATATACATCAACATTTACACATACATATATATATA 16046
OY 1634 TATTAAGAAAGTCAAGCTTTTAAATAATCTGCATTTGATGATGTTTAAAGTAGATCA 1693
    || || || || || || || || || || || || || || || || || || || || ||
Db 16047 TATATATACCCATTAATCAATACATTTTACATACATATATATATATATATAT 16106
OY 1694 AAA 1696
    || ||
Db 16107 ATA 16109

```

RESULT 2  
US-09-641-638-651/C  
: Sequence 651, Application US/09641638  
: Patent No. 6432648  
: GENERAL INFORMATION:  
: APPLICANT: Blumenfeld, Marta  
: APPLICANT: Bougueleret, Lydie  
: APPLICANT: Chumakov, Ilya  
: APPLICANT: Cohen, Annick  
: TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
: FILE REFERENCE: GENSET, 051CPI  
: CURRENT FILING DATE: 2000-08-16  
: PRIOR FILING DATE: 2000-02-11  
: PRIOR APPLICATION NUMBER: US 09/502,330  
: PRIOR FILING DATE: 2000-02-11  
: PRIOR APPLICATION NUMBER: US 60/133,200  
: PRIOR FILING DATE: 1999-05-07  
: PRIOR APPLICATION NUMBER: US 09/275,267  
: PRIOR FILING DATE: 1999-03-23  
: PRIOR APPLICATION NUMBER: US 60/119,917  
: PRIOR FILING DATE: 1999-02-12

NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: Patent.pm  
SEQ ID NO 651  
LENGTH: 20674  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1123..3123  
OTHER INFORMATION: 5'regulatory region  
NAME/KEY: exon  
LOCATION: 3124..3297  
OTHER INFORMATION: exon 1  
NAME/KEY: exon  
LOCATION: 3871..4072  
OTHER INFORMATION: exon 2  
NAME/KEY: exon  
LOCATION: 5552..5633  
OTHER INFORMATION: exon 3  
NAME/KEY: exon  
LOCATION: 5758..5880  
OTHER INFORMATION: exon 4  
NAME/KEY: exon  
LOCATION: 5996..6099  
OTHER INFORMATION: exon 5  
NAME/KEY: exon  
LOCATION: 6349..6509  
OTHER INFORMATION: exon 6  
NAME/KEY: exon  
LOCATION: 7379..7522  
OTHER INFORMATION: exon 7  
NAME/KEY: exon  
LOCATION: 8645..8854  
OTHER INFORMATION: exon 8  
NAME/KEY: exon  
LOCATION: 12254..12340  
OTHER INFORMATION: exon 9  
NAME/KEY: exon  
LOCATION: 12854..13023  
OTHER INFORMATION: exon 10  
NAME/KEY: exon  
LOCATION: 13308..13429  
OTHER INFORMATION: exon 11  
NAME/KEY: exon  
LOCATION: 16567..16667  
OTHER INFORMATION: exon 12  
NAME/KEY: exon  
LOCATION: 16775..16945  
OTHER INFORMATION: exon 13  
NAME/KEY: exon  
LOCATION: 17063..17554  
OTHER INFORMATION: exon 14  
NAME/KEY: misc\_feature  
LOCATION: 17555..20674  
OTHER INFORMATION: 3'regulatory region  
NAME/KEY: allele  
LOCATION: 1128  
OTHER INFORMATION: 10-508-191 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1182  
OTHER INFORMATION: 10-508-245 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1559  
OTHER INFORMATION: 10-509-284 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1570  
OTHER INFORMATION: 10-509-295 : deletion of C  
NAME/KEY: allele  
LOCATION: 1827  
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTT  
NAME/KEY: allele  
LOCATION: 2048  
OTHER INFORMATION: 10-511-62 : polymorphic base C or T

NAME/KEY: allele  
LOCATION: 2323  
OTHER INFORMATION: 10-511-337 : insertion of T  
NAME/KEY: allele  
LOCATION: 2341  
OTHER INFORMATION: 10-512-36 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 2623  
OTHER INFORMATION: 10-512-318 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2832  
OTHER INFORMATION: 10-513-250 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2844  
OTHER INFORMATION: 10-513-262 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 2934  
OTHER INFORMATION: 10-513-352 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2947  
OTHER INFORMATION: 10-513-365 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 3602  
OTHER INFORMATION: 12-206-81 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 4062  
OTHER INFORMATION: 10-343-231 : deletion of C  
NAME/KEY: allele  
LOCATION: 4088  
OTHER INFORMATION: 12-206-366 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4109  
OTHER INFORMATION: 10-343-278 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4170  
OTHER INFORMATION: 10-343-339 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 5903  
OTHER INFORMATION: 10-346-23 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6019  
OTHER INFORMATION: 10-346-141 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6141  
OTHER INFORMATION: 10-346-263 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6183  
OTHER INFORMATION: 10-346-305 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6338  
OTHER INFORMATION: 10-347-74 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6375  
OTHER INFORMATION: 10-347-111 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6429  
OTHER INFORMATION: 10-347-165 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6467  
OTHER INFORMATION: 10-347-203 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6484  
OTHER INFORMATION: 10-347-220 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6534  
OTHER INFORMATION: 10-347-271 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 6611  
OTHER INFORMATION: 10-347-348 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 7668  
OTHER INFORMATION: 10-348-391 : polymorphic base A or G  
NAME/KEY: allele

LOCATION: 8608  
OTHER INFORMATION: 10-349-47 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 8658  
OTHER INFORMATION: 10-349-97 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8703  
OTHER INFORMATION: 10-349-142 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 8777  
OTHER INFORMATION: 10-349-216 : deletion of CTG  
NAME/KEY: allele  
LOCATION: 8785  
OTHER INFORMATION: 10-349-224 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 8926  
OTHER INFORMATION: 10-349-368 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12171  
OTHER INFORMATION: 10-350-72 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12429  
OTHER INFORMATION: 10-350-332 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13341  
OTHER INFORMATION: 10-507-170 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 13492  
OTHER INFORMATION: 10-507-321 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 13524  
OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13535

Query Match 3.68; Score 65.8; DB 4; Length 20674;  
Best Local Similarity 46.68; Pred. No. 0.00011;  
Matches 245; Conservative 0; Mismatches 279; Indels 2; Gaps 1;

QY 1078 AACGAGTAAACAAATACCTTAATATATATCTGATATATATTTGACGCTGATATTT 1137  
DB 11584 AGCTATATGAAAAAGCCATGCTCTTTTATATATTAATAATCTTACCTATTA 11525  
QY 1138 AATATATTAAGATATATCTATATAGCTTAATATATACCTAAAGAGTAATATAG 1197  
DB 11524 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11465  
QY 1198 GATTGCTATTTAAAAAGCATTGAGAAATGAAATGAGACATTATAAAAAGCGTGAT 1257  
DB 11464 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11405  
QY 1258 ATCAAAATCTTTTACAAATCTGCAATTAATTAATTAATTAATTAATTAATTAATTA 1317  
DB 11404 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11345  
QY 1318 GAATGCGCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1377  
DB 11344 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11285  
QY 1378 GTTGAGACAAATTTGCGCTGCTTCAATTAATTAATTAATTAATTAATTAATTAATTA 1437  
DB 11284 GTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11227  
QY 1438 TCATCAGCTAAACGTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1497  
DB 11226 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11167  
QY 1498 TGGGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1557  
DB 11166 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11107  
QY 1558 GATAATTAACATATGAGAAATTAACATTAATTAATTAATTAATTAATTAATTAATTA 1603



DB 980 CAGATGGTCTTTTATCATCTATGTAATCCCAAGATACATTATTTAAANA 1039  
1785 TCACATTCG 1794  
DB 1040 TGAGACCTAG 1049

## RESULT 5

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEFFELINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PTZgpt-Fls

US-08-232-463-14

Query Match 3.4%; Score 63.2; DB 1; length 7218;

Best Local Similarity 16.0%; Pred. No. 0.00031;

Matches 91; Conservative 224; Mismatches 252; Indels 1; Gaps 1;

QY 1277 ACTGCATTACTTGTATAGTAATGATGATCAATGTCGCGATTTATANA 1336  
DB 1643 ACCTGACGAATCTTATATAGCATGAAAAAATGAATTTACTATATTTTAA 1584  
QY 1337 ACTCAATTTATCCATGATGAAAGTTTGAATCTTGTAGTGAGACAATTCGGCT 1396  
DB 1583 CATACATATATCTTAATATGAAGTGGATGTGACTAGCGTACGCTTCTAGAC 1524  
QY 1397 GGGCTTTCAATTAATATACAGACGAATGAATTTATTCACAGCTAAACTGAA 1456  
DB 1523 GCATCTATTCAGTTTCAAAAAAGCGCATGTAGCATCACTGAATTACTATCTATGCA 1464

QY 1457 ATGATTTTAAAAAATACGCTTCTTATTCATTCGATTCGAGAAATTAATACG 1516  
DB 1463 GT-AGTTAAAGATAGAAGATTGGTACRRRRRRRRRRRRRRRRRRRRRRRR 1405  
QY 1517 ATGAGATGATCAATATACATCAAGAAATGATTAAGAGATTAGATAATACATATGAG 1576  
DB 1404 RRR 1345  
QY 1577 AATAACCATATGATTAATTAAGAGATATATGCTTTGCGAAATACAAAGATAT 1636  
DB 1344 RRR 1285  
QY 1637 AAAAGAACTGACCTTAAATATCTCATTCATGATGCTTTTAAAGATGAGAAA 1696  
DB 1284 RRR 1225  
QY 1697 ATTTATGATGAAGAGATTTGCACTAAGATATGAAATATGATGAAATCTTAT 1756  
DB 1224 RRR 1165  
QY 1757 TGATGATCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1816  
DB 1164 RRR 1105  
QY 1817 CCGCATTAATAAACAGCAAGCAAAAGAGA 1844  
DB 1104 RRR 1077

## RESULT 6

US-08-487-826B-13/C

Sequence 13, Application US/08487826B

Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitulis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobe Martens Olson &amp; Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,826B

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelien, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CPI

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA





;; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
;; NUMBER OF SEQUENCES: 39  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BERESKIN & PARR  
;; STREET: 40 King Street West  
;; CITY: Toronto  
;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: M5H 3Y2  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/883,795A  
;; FILING DATE: 27-JUN-1997  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gravelle, Micheline  
;; REGISTRATION NUMBER: 40,261  
;; REFERENCE/DOCKET NUMBER: 7841-062  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 364-7311  
;; TELEFAX: (416) 361-1398  
;;  
;; INFORMATION FOR SEQ ID NO: 36:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 665 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cdna  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; IMMEDIATE SOURCE:  
;; CLONE: Rh 32  
;; US-08-883-795A-36

Query Match 3.38; Score 61.6; DB 2; Length 665;  
Best Local Similarity 46.7%; Pred. No. 0.00043;  
Matches 196; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 926 ACAAAATTAATAAAGAAAGAAATTTGAGCGAAGAAATTTTGAATTTTAAAG 985  
DB 447 ATAAATATGTATTAATTAACATTTTAATTAATAATATGTATTAATTAACATTTTAATT 388  
QY 986 GCGATCTTGCTACCGCATTTTGGCATTTTAAACCTGACATCTTTTAAGTTAATA 1045  
DB 387 ATAAATATGTATTAATTAACATTTTAATTAATAATATGTATTAATTAACATTTTAATT 328  
QY 1046 GATATATCGGTAGATTATAAGTATTAATAAAGCAATAAACAATTAATTAATT 1105  
DB 327 ATAAATATGTATTAATTAACATTTTAAATTAATAATTAATTAATAACATTTTAATT 268  
QY 1106 TAAATTCGATTAATTTGACGATTAATTAATTAATAAGATATATTCATTTGCT 1165  
DB 267 ATAAATATTTTAATTAATTAATTTTAATTAATAATTAATTAATTAATAATTTTAATT 208  
QY 1166 TAAATTAATTAATAAAGAGTAATAATATGTATTTGATTTTAAAGCAATTAAGAA 1225  
DB 207 ATAAATATTTTAATTAATTAATTTTAATTAATAAATTAATTAATAATTTTAATT 148  
QY 1226 ATGAATAGACATTTATAAAGAGCGGTATATCAATCTTTCTTAACAATCTTGCAATT 1285  
DB 147 ATAAATATTTTAATTAATAATTTTAATTAATAAATTAATTAATAATTTTAATT 88  
QY 1286 ACTTGATATAGATTAAGCAATTAAGTGAATGAGATGCAATTAATTAATTAACCAAT 1345  
DB 87 ATAAATATTTTAATTAATAATTTTAATTAATAATTTTAATTAATAATTTTAATT 28

RESULT 9  
US-08-998-416-186/c

;; Sequence 186, Application US/08998416  
;; Patent No. 6239264  
;;  
;; GENERAL INFORMATION:  
;; APPLICANT: Philippsen, Peter  
;; APPLICANT: Pohlmann, Rainer  
;; APPLICANT: Steiner, Sabine  
;; APPLICANT: Mohr, Christine  
;; APPLICANT: Wendland, Jurgen  
;; APPLICANT: Knechtle, Philipp  
;; APPLICANT: Reibischung, Corinne  
;; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPYII  
;; NUMBER OF SEQUENCES: 1152  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: No. 6239264atlis Corporation  
;; STREET: 3054 Cornwallis Road  
;; CITY: Research Triangle Park  
;; STATE: No. 6239264th Carolina  
;; COUNTRY: USA  
;; ZIP: 27709  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/998,416  
;; FILING DATE: 24-DEC-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: CH 0016/97  
;; FILING DATE: 31-DEC-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Weigs, J. Timothy  
;; REGISTRATION NUMBER: 38,241  
;; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 919-541-8587  
;; TELEFAX: 919-541-8689  
;;  
;; INFORMATION FOR SEQ ID NO: 186:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 615 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: PAG1074RP  
;; US-08-998-416-186

Query Match 3.38; Score 61.4; DB 4; Length 615;  
Best Local Similarity 46.8%; Pred. No. 0.00047;  
Matches 269; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

QY 1061 TTATTAAGTATGTAAAGAAAGAGTAATAAACAATTAATTAATTTTGAATTTAATA 1120  
DB 608 TTATTAATATTTTATTAATTAATTAATAATTAAGATTAATAATTAATTAATTAATTA 549  
QY 1121 TTGACAGTGATTTATTAATATTAATAAGATATATCTATTAAGCTTAATAATTAATAA 1180  
DB 548 TT--AAATTAATAATAATAATAATAATGAATATTAATTTTAATAATTAATAATAATAA 491  
QY 1181 AAAGAGTAATATATGATTTGCTATTTAAAGCAATTAAGAAATGAATTAAGACATT 1240  
DB 490 TAAAGAAATTAAGTAAAGTAAATTTTAATTAATTAATTAATTAATAAAGTTAAATAATA 431  
QY 1241 ATAAAGAGCGGATATCAATCTTTCTTAACAATCTTGCAATTAATTTGATTTAGATA 1300  
DB 430 AATTAACATTAATTTTAAAGAAATGAATATTAATTAATAAATTAATTAATTAATTTTA 371  
QY 1301 AAGCATTAATGATGATGATGATGCGATATTAATAAATTAATTAATTAATTAATTAATTA 1360  
DB 370 AATAAATTAATCTTTTAATTAATTAATAATTAATTAATTAATTAATTAATAAATTAATAAT 311

```

OY 1361 GTTTGATCTCTTGATGTGACACAAATTCGGCTTTCATATAAATATACAGA 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 AATATTAAT-----TATTGATTAATCTTTAATTAATTAATTAAGAAATAATATAT 259
OY 1421 GACGAATGCAATATTTATCATGCTAAACGAAATGATTTTAAAAATCTCGTCT 1480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 258 CTAAATAATTTTAACTAACTAAATTTGAACATGACTAAATAGTATTCATATTA 199
OY 1481 TTATTCATCGATTCGATGGAAGAAATTAATAGATGAGATGATACATCATCAG 1540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 AATATTTATTTATTTATTTATTAATAATTAATTAATGATCAATTAATTAATTA 139
OY 1541 AATGATTAAGCATTTAGATTAATTCACATATGAGAAATTAACATGATTAATTAAT 1600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 138 AATATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 79
OY 1601 AAGAAATATTTGATTTGCAATTAACAAAGATA 1635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78 TAAATTAATTTAACTTAATTAATTTAATTAATTAATTAATTAATTAATTAAT 44

```

# RESULT 10

US-07-867-106-2/c

Sequence 2, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved Plasmid Vectors for Cellular

TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Peeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5852 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

## US-07-867-106-2

Query Match 3.3%; Score 60.2; DB 1; Length 5852;

Best Local Similarity 44.5%; Pred. No. 0.0012;

Matches 326; Conservative 0; Mismatches 403; Indels 4; Gaps 2;

```

OY 1018 AAACCTGACATCTTTATTAAGTAAATAGATATATCCGTAGATTAATAGATGTTAA 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2408 AAACCTATCCAGAAATAGATGCTGCATTTTAAATCTTGAATTAATTAACCA 2349
OY 1078 AACGATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2348 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2289
OY 1138 AATATATTAAGATATATCTATTAATTAATTAATTAATTAATTAATTAATTA 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2288 AAGTGTGGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2230
OY 1198 GATTGTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2229 AAAAAAAAAAAAAAAAAAAATTAATCAAAATTAATTAATTAATTAATTAATTA 2170
OY 1258 ATCAATCTTTCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAAT 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2169 ACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2110
OY 1318 GATGTCGCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2109 TAAATTTTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2050
OY 1378 GTTGACACAAATTCGGCTGCGCTTCAATTAATTAATTAATTAATTAATTAAT 1437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2049 TGAGGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1990
OY 1438 TCATGACGTAACGAAATGATTTTAAATTAATTAATTAATTAATTAATTAATTA 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1989 AAAAAAGAAATGAAGAAATGTTGTTAATCTACATTAATTTTATTAATTTTGCAT 1930
OY 1498 TCGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1929 TTAATAATTAATCTTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 1873
OY 1558 GATTAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1872 AAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1813
OY 1618 TCGAATTAACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1812 AAAATTTTATTTCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTA 1753
OY 1678 TTTTAAAGTAGAGAAATTTTATTTGATGAAGAGATTTGCACTAAGATATTGA 1737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1752 TATCTAATAAAATTAACAGAAATTAATTAATTAATTAATTAATTAATTAATTA 1693
OY 1738 TATCTATGAGAAA 1750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1692 TTTTTTTAAAAA 1680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 11

US-08-998-416-1137/c

Sequence 1137, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtel, Philipp

APPLICANT: Reibschung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII

TITLE OF INVENTION: AND USES THEREOF

QY	1013	TATTTAAACCTGACACTCTTTATAGTAATACATATATCCGTTAGTTTAAAGTATG	1072
Db	636	TTTTTAAGAATTTTTTAACTACATCTTTTAAATATTTTATTAATTAATTAATATGAT	577
QY	1073	TTAAAAACGACTAAAAACAATACCTTATATATTTAAATCTCAATATATTTTGACAGTAT	1132
Db	576	AAATATTTAATTAATATATTAATTTAATTAATTAATTAATTAATTAATTAATTAAT	517
QY	1133	TATTTAATATTTTAAGAGATATATCTATTTACCTTAATATATACATAAAAAAGAGTAAAT	1192
Db	516	TATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAATTAATTA	457
QY	1193	ATATGATCTGCTATTTTAAAAACATTAAGAAATGAACATATATTAATAAGACG	1252
Db	456	TTCTTATTAATAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	400
QY	1253	GTGATATCAAAATCTTCTTACATACCTTGCAATCTGATGATATAGATAAAGCATTAATG	1312
Db	399	ATATATTAATAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	340
QY	1313	GTGATGATGCGGGGATTTATTAATTAACATTAATTAATCATGATGAAGAAGTTTGAAT--CT	1370
Db	339	ATTATTTTAAATTAACATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	280
QY	1371	TCTTGATGTGGAGACAATTTTCGGCTGGCTTTCAATTAATAATTAATTAACAGACAGCAATGA	1430
Db	279	TATTAAGAAATATATATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	220
QY	1431	ATATTTATCACTACACTAAACTGAAGAATGATTTTAAAAAATATCTGTTCTTATATTCATTC	1490
Db	219	GACATTAATATAGT--ATTCATATTTAAATATATTAATTTATTAATTAATTAATTAATTAATTAAT	162

2030 AGAATAAGAAACCT  
1457 ATGATTTTAAAAAA

[illegible]

Db	1970	TTATATATATATATATATATAGTGAGGTTTTTCTACTCTCATTTTTTTTGGAACATTTTAAAGATTT	1911
QY	1517	ATGAGATGAGTACATACATCATCAAGAAATGATTTAAAGCATTTGATATATACACATATGGAG	1576
Db	1910	CGGATTTATCAAAATATATGTTATTTCTAGAAATTAATATATATATCTCTAAATATATATATA	1851
QY	1577	AAATPACCATCTGATATATAAAATPAGAAATATATTCGATTTTCGAAATPACAAAGATAT	1636
Db	1850	TTGTATAGCATATCAT	1791
QY	1637	AAAAAGAAAGTCAGCTTTAAAAATATCTGCATTGATCGATGTTTTTAAAACTAGATGAAA	1696
Db	1790	TAGAAATPACAAAGGATATAGTTATTCGTGACACAAAAAAATTTTAAACCGAAAA - AAAA	1733
QY	1697	ATTATTTGATCAAGAGATTTGCCAATPACATATTTGAAAATATTCGTATGAAAAATCCTAT	1756
Db	1732	ATTAATATATTTAAAAAATGTGATCTCACTAAATATATATATATATATATATATATATATATATAT	1673
QY	1757	TGATGATCCAGATGGCATPAGAAAT	1785
Db	1672	AAAAAAAAAAAAATTTGATATTTATATATAT	1644

ORGANISM: PAG1241RP  
US-08-998-416-288

Query Match	3.18;	Score 58;	DB 4;	Length 837;
Post [cont.] Similarity	45.29;	Prod NO	0.0024.	

[illegible]

FILING DATE: 06-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29,009  
REFERENCE/DOCKET NUMBER: 47-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8920 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic  
US-08-446-855A-1

Query Match 3.1%; Score 58; DB 2; Length 8920;  
Best Local Similarity 46.8%; Pred. No. 0.0035;  
Matches 249; Conservative 0; Mismatches 280; Indels 3; Gaps 2;

QY 923 AAAACAAAAAATAAACAAGAAAAATTTTCGAGCGAAAAATATTTTGGAAATTTTAA 982  
DB 563 AAATTAT 504  
QY 983 AAGCGATCTCTGCTACCGCCTTTTCCATATTTAAACCTGACTATCTTTATAGTTA 1042  
DB 503 TATTTGTAT 444  
QY 1043 ATGATATATATCCGTAT 1102  
DB 443 ACACATCAAAAT 384  
QY 1103 ATTATATCTGAT 1162  
DB 383 ATTTCAAGTAAATTTTATTA-ACAAATGACACACAACAATATATATATATATATAT 325  
QY 1163 GCTTAATATATCAATAAAGAGGTAATATATATATATATATATATATATATATAT 1222  
DB 324 AT 265  
QY 1223 AAAATGAATATAGAACATTTATTAAGAGCGTATATCAATCTTTCTTCAATATCTGC 1282  
DB 264 AAAAAAAGAGTTTAT 205  
QY 1283 ATTACTTGTAT 1342  
DB 204 ATATCAAAAT 145  
QY 1343 AT--TATCATTTGATGAAAGTTTGTCTCTGTGATGTGAGCACAATTTGCGCTGGGC 1400  
DB 144 ATGCTTATATAGATTTTACTCTTTCTCTGTAAAGCTTTTAAAGTCCCTTCAAGTATCG 85  
QY 1401 TTTCAATTAATAATATACAGAGCAAAATATATATATATATATATATATATATAT 1452  
DB 84 AATTTCTAACCTCATTCACAAAGAGAGTGTGTTTCTTTCTTTAAATTT 33

RESULT 15  
US-09-150-741-1/c  
Sequence 1, Application US/09150741  
Patent No. 6183996  
GENERAL INFORMATION:  
APPLICANT: Stewart et al.  
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
Patent No. 6183996  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/09/150,741  
CURRENT FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: PL6380  
EARLIER FILING DATE: 1992-12-16

EARLIER APPLICATION NUMBER: AU93/00617  
EARLIER FILING DATE: 1993-12-02  
EARLIER APPLICATION NUMBER: 08/446,855  
EARLIER FILING DATE: 1995-07-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: Patentl Ver. 2.0  
SEQ ID NO 1  
LENGTH: 8920  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
US-09-150-741-1

Query Match 3.1%; Score 58; DB 4; Length 8920;  
Best Local Similarity 46.8%; Pred. No. 0.0035;  
Matches 249; Conservative 0; Mismatches 280; Indels 3; Gaps 2;

QY 923 AAAACAAAAAATAAACAAGAAAAATTTTCGAGCGAAAAATATTTTGGAAATTTTAA 982  
DB 563 AAATTAT 504  
QY 983 AAGCGATCTCTGCTACCGCCTTTTCCATATTTAAACCTGACTATCTTTATAGTTA 1042  
DB 503 TATTTGTAT 444  
QY 1043 ATGATATATCCGTAT 1102  
DB 443 ACACATCAAAAT 384  
QY 1103 ATTATATCTGAT 1162  
DB 383 ATTTCAAGTAAATTTTATTA-ACAAATGACACACAACAATATATATATATATATAT 325  
QY 1163 GCTTAATATATCAATAAAGAGGTAATATATATATATATATATATATATATATAT 1222  
DB 324 AT 265  
QY 1223 AAAATGAATATAGAACATTTATTAAGAGCGTATATCAATCTTTCTTCAATATCTGC 1282  
DB 264 AAAAAAAGAGTTTAT 205  
QY 1283 ATTACTTGTAT 1342  
DB 204 ATATCAAAAT 145  
QY 1343 AT--TATCATTTGATGAAAGTTTGTCTCTGTGATGTGAGCACAATTTGCGCTGGGC 1400  
DB 144 ATGCTTATATAGATTTTACTCTTTCTCTGTAAAGCTTTTAAAGTCCCTTCAAGTATCG 85  
QY 1401 TTTCAATTAATAATATACAGAGCAAAATATATATATATATATATATATATATAT 1452  
DB 84 AATTTCTAACCTCATTCACAAAGAGAGTGTGTTTCTTTCTTTAAATTT 33

Search completed: July 17, 2003, 05:12:46  
Job time : 94.0057 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 19:52:25 ; Search time 4292.71 Seconds

(without alignments)  
11254.112 Million cell updates/sec

Title: US-08-876-132-1

Perfect score: 1660

Sequence: 1 AGATCTACACAGCAAAATT.....AAGACTCCGACGACATCT.1660

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

Genembl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vl:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vl:\*  
30: em.htg.hum:\*  
31: em.htg.in:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.rod:\*  
36: em.htg.mam:\*  
37: em.htg.vrt:\*  
38: em.sv:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1308.6	78.8	3661	1	AF128889	AF128889 Pantoea C
2	109	6.6	6282	6	AX251493	AX251493 Sequence
3	106.2	6.4	108908	3	PFMAL3P8	AL034560 Plasmid
4	104.2	6.3	349980	6	AX344555	AX344555 Sequence
5	101.6	6.1	131682	9	AL672277	AL672277 Human DNA
6	101.6	6.1	231912	2	AC087566	AC087566 Mus muscu
7	99.6	6.0	174384	2	AC009524	AC009524 Homo sapi
8	98.4	5.9	130540	2	AC079417	AC079417 Mus muscu
9	96.6	5.8	14867	3	AE001398	AE001398 Plasmid
10	96.6	5.8	84563	5	AC096885	AC096885 Danio rer
11	93.6	5.6	6644	6	E23356	E23356 Virus vecto
12	93.6	5.6	7372	6	E23357	E23357 Virus vecto
13	93.6	5.6	7797	6	E23359	E23359 Virus vecto
14	93.6	5.6	268147	2	AC116966	AC116966 Dictyoste
15	93.4	5.5	67970	3	PFMAL1P3	AL034104 Plasmid
16	91.8	5.5	191366	2	AC115351	AC115351 Rattus no
17	91.4	5.5	257703	3	CEY111B2A	AL132904 Caenorhab
18	90.8	5.4	14006	6	AX346860	AX346860 Sequence
19	90.2	5.4	11000	2	CEY111B2_0	Z98857 Caenorhabd
20	90	5.4	300695	2	AC079431	AC079431 Mus muscu
21	89.8	5.4	6668	6	AX346599	AX346599 Sequence
22	89.8	5.4	16544	2	AC125605	AC125605 Homo sapi
23	88.4	5.3	83234	2	AC034104	AC034104 Homo sapi
24	88.4	5.3	5420	3	AF056936	AF056936 Plasmid
25	87.6	5.3	7442	6	AX251441	AX251441 Sequence
26	87.6	5.2	170568	2	AC107605	AC107605 Rattus no
27	86.6	5.2	15548	6	AX347057	AX347057 Sequence
28	86.4	5.2	303091	2	AC084799	AC084799 Mus muscu
29	86.2	5.2	156060	2	AC004153	AC004153 Plasmid
30	86.2	5.2	169794	2	AC004688	AC004688 Plasmid
31	86.2	5.2	224448	2	PFMAL4P4	AL035477 Plasmid
32	86	5.1	9716	2	AC115603	AC115603 Dictyoste
33	85.4	5.1	66441	3	PFMAL1P4	AL031747 Plasmid
34	85	5.1	215287	2	AC079432	AC079432 Mus muscu
35	85	5.1	147962	2	AC130632	AC130632 Rattus no
36	84.6	5.1	82469	2	AC116330	AC116330 Dictyoste
37	84.2	5.1	143585	2	AC013349	AC013349 Homo sapi
38	84	5.0	9539	6	AX277889	AX277889 Sequence
39	83.8	5.0	9539	6	AX323566	AX323566 Sequence
40	83.6	5.0	7571	6	AX345429	AX345429 Sequence
41	83.6	5.0	7823	6	AX278035	AX278035 Sequence
42	83.6	5.0	7823	6	AX323804	AX323804 Sequence
43	83.6	5.0	7823	6	AX344489	AX344489 Sequence
44	83.6	5.0	7823	6	AX346963	AX346963 Sequence
45	83.6	5.0	7823	6	AX346963	AX346963 Sequence

#### ALIGNMENTS

RESULT 1  
AF128889  
LOCUS  
DEFINITION Pantoea citrea strain ATCC31623 plasmid pPZG500, complete plasmid  
ACCESSION AF128889  
VERSION AF128889.1 GI:44572229  
KEYWORDS  
SOURCE Pantoea citrea.  
ORGANISM Pantoea citrea.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
REFERENCE 1 (bases 1 to 3661)  
AUTHORS Billie M. and Delic V.  
TITLE Isolation and characterization of a cryptic plasmid from *Erwinia*

JOURNAL citreus ATCC 31623  
MEDLINE J. Appl. Microbiol. 83 (4), 485-492 (1997)  
PUBMED 98012514  
REFERENCE 9351229  
AUTHORS 2 (bases 1 to 3661)  
TITLE Bilic Nezelic M. and Dellec V.  
JOURNAL Sequence Analysis and Functional Characterization of Replication  
REFERENCE Region in Plasmid pPZG500 from Pantoea citrea ATCC 31623  
AUTHORS Unpublished  
JOURNAL 3 (bases 1 to 3661)  
REFERENCE Bilic Nezelic M. and Dellec V.  
AUTHORS Direct Submission  
JOURNAL Submitted (16-FEB-1999) Biotechnology, PLIVA d.d., Research  
TITLE Institute, Pilaz baruna Filipovica 25, Zagreb 10 000, Croatia  
FEATURES Location/Qualifiers  
source 1. 3661  
/organism="Pantoea citrea"  
/strain="ATCC31623"  
/db\_xref="ATCC:31623"  
/db\_xref="taxon:53336"  
/plasmid="pPZG500"  
complement(join(3188..3661,1..177))  
/note="RNAI"  
99..317  
/note="Orf1"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAD21205.1"  
/db\_xref="GI:4457230"  
/translation="MGQALCWSAHPIYSIGALAKLKHKNLIGLGLYOKKKIK  
OKRIFERNILFEFGDTCYRFALFKT"  
177  
/note="Orf4"  
complement(334..342)  
/note="putative"  
/bound\_moiety="DnaA"  
458..472  
/bound\_moiety="IHP"  
489..704  
/note="Orf2"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAD21206.1"  
/db\_xref="GI:4457231"  
/translation="MDCVFKKALENEIEHYKKDIDKSFQYLHYFDIDKALNGDEG  
DIINSNTSIDSLIFLMLSSAGLSIK"  
complement(845..1066)  
/note="Orf3"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAD21207.1"  
/db\_xref="GI:4457232"  
/translation="MPSGSSIGFSYDIFNIFSCSSSSINFSSTLKHRSQIIFLRAD  
FLYLILFRNOYLILFYISGYFSICVIL"  
complement(1193..1348)  
/note="Orf4"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAD21208.1"  
/db\_xref="GI:4457233"  
/translation="MODLCWHLFFLHPDIFYTHHFLFLYLKFSMMKFLSTERAENI  
SHIFLKK"  
1991..2365  
/note="Orf5"  
/codon\_start=1  
/transl\_table=11  
/product="putative heat-shock protein"  
/protein\_id="AAD21209.1"  
/db\_xref="GI:4457234"

CDS  
/translation="MDKSRKLTLEKKDLEKIYAESEIKAKKLTGPGVLEMTKE  
MKKNINLVNETAGQYRKLFENKRVKHSKSDLVLTGLECGTRNSFDTRSAFRCIC  
ERIQWNDIEFTTKILQYKEKQ"  
2362..2991  
/function="putative resolution of catenane plasmid dimers  
prior to plasmid segregation to the daughter cells"  
/note="Orf6"  
/codon\_start=1  
/transl\_table=11  
/product="lambda-integrase like protein"  
/protein\_id="AAD21210.1"  
/db\_xref="GI:4457235"  
/translation="WKEADTMDIDFKRLKNNKSYDRAGLSTICSTGCPRAEVLK  
LEIYNRYEDISFKIIGAKVNDGQSERLHFDLSKYHNDQMNVTLSOLKQKF  
YKPDGKLYNSKROYLYIOHRTFSLYLHRRAVDLAKSGADFTTAFLGHVTOSE  
FVGVARSSKGGIAVTGVESCDVYKANKSQFAVSRPSQISTLSHKFK"  
-35.signal 3154..3159  
-10.signal 3175..3180  
misc\_RNA 3188..3288  
/note="RNAI"  
-10.signal 3295..3300  
-35.signal 3311..3316  
BASE COUNT 1366 a 472 c 730 g 1093 t  
ORIGIN  
Query Match 78.8%; Score 1308.6; DB 1; Length 3661;  
Best Local Similarity 91.1%; Pred. No. 5e-182;  
Matches 1510; Conservative 0; Mismatches 12; Indels 136; Gaps 6;  
8 CACAAGCGAATTGAAAAATAGATAAAA-TTTTGGCAGGRTTTAAACCCGACTTAAAC 66  
1433 CACAAGCGAATTGAAAAATAGATAAAA-TTTTGGCAGGRTTTAAACCCGACTTAAAC 1492  
67 AATGAGTGAAGAAGAAAAAATAATATACA-TATTTGAGTAGTAAAGAGAA 124  
1493 AATGAGGAGAGAGAAAAAATTAATACCTTATTGGAGTTAGTAAAGAGAA 1552  
125 AGAAAAATTAAGAAGACCTCGCTTAACAGTCGAAAAACCGAATTAATAAAAAGAA 184  
1553 AGAAAAATTAAGAAGACA-CTCGGCTTAACAGTCGAAAAACCGAATTAATAAAAAGAA 1611  
185 GAGACTGATTTTAAATGGAATGCTGAGGAAAAAATTTAATTTTCAATTTTCGAG 244  
1612 GAGACTGATTTTAAATGGAATGCTGAGGAAAAAATTTAATTTTCAATTTTCGAG 1671  
245 GGATTAAATTTGTTGAAGTGTGATAAATCTAGATAAATAATGAGATCAAAAATGTGT 304  
1672 GGATTAAATTTGTTGAAGTGTGATAAATCTAGATAAATAATGAGATCAAAAATGTGT 1731  
305 TGAATTTGACATTTATGAATACGTAGTATATCAATTAATGGGGTTGTCTATTTATTT 364  
1732 TGAATTTGACATTTATGAATACGTAGTATATCAATTAATGGGGTTGTCTATTTATTT 1791  
365 TGGCAAGTGAATAATCGAGGAGAAAAATAGTTTGGAGAGAAAAAAACCTTGCC 424  
1792 TGGCAAGTGAATAATCGAGGAGAAAAATAGTTTGGAGAGAAAAAAACCTTGCC 1851  
425 GTTTTTCGAATGACTTTGAAAAAATTCATTGTGACGGTAGCGAACTTTGAATTT 484  
1852 GTTTTTCGAATGACTTTGAAAAAATTCATTGTGACGGTAGCGAACTTTGAATTT 1911  
485 TTTACATTGGAATTTGAAAAAATTAAGCAAAAGAACTCAAAATGGAATTTATTA 544  
1912 TTTACATTGGAATTTGAAAAAATTAAGCAAAAGAACTCAAAATGGAATTTATTA 1971  
545 TAAAAAAGAGATCGGATATGATTTAAAGCAAAAGAACTCAATGATGAAATTA 604  
1972 TAAAAAAGAGATCGGATATGATTTAAAGCAAAAGAACTCAATGATGAAATTA 2031  
605 AGATTGGAAAAAATCTATGCTGAGAGTGAATTAAGCAAAATTTGGAACTCAAC 664  
2032 AGATTGGAAAAAATCTATGCTGAGAGTGAATTAAGCAAAATTTGGAACTCAAC 2091



OY		665	CGGGTGTCTTTAGCAATGACGATGAAGAAGAAATGAAAATATATCATCCGCATGTAA	724
Db		2092	CGGGTGTTTGTAGCAATGACGATGAAGAAGAAATGATAAATATATCATCCGCATGTAA	2151
OY		725	TGAAGAAACACACAGGTCATATATAGCAATATTATCCAAAAATTAAGTGTGACATAGTAATC	784
Db		2152	TGAAGAAACACACAGGTCATATATAGCAATATTATCCAAAAATTAAGTGTGACATAGTAATC	2211
OY		785	AGATGATCTAGTAACGGGACTATATTAGACTGTGAACTCGAAATAGTTTTGATPAAAAACAAG	844
Db		2212	AGATGATCTAGTAACGGGACTATATTAGACTGTGAACTCGAAATAGTTTTGATPAAAAACAAG	2271
OY		845	AAGGCCCTTCGTTTTTGTATTGTGTGAGAGAAATTCAGCACTGAGAAAAAGACGTATAA	904
Db		2272	AAGGCCCTTCGTTTTTGTATTGTGTGAGAGAAATTCAGCACTGAGAAAAAGACGTATAA	2299
OY		905	TGCAAGAAAGATAAAGATTCGATACAAATGAAGCAAAAACATAAGAGCTTTGANTT	964
Db		2300	-----	2299
OY		965	GAGTTTGTCTTTTGATTAAGATTTTTTGGAGTGAAAATAGAAATTCATAGGAATGATATT-	1023
Db		2300	-----AGATTCATAGGAATGATATT	2322
OY		1024	-CTCACACACAAAAAAGACTCTGCAGATTAAGAAAAAACATGAAGAAAGCGGACACAAATG	1082
Db		2323	CTTCACACACAAAAAAGATCTGCAGATTAAGAAAAAACATGAAGAAAGCGGACACAAATG	2382
OY		1083	GATGATATATTTTAAGAGGCTPAAAAAATATATTAATCTACATATATGATTCGTTATGCTGATTC	1142
Db		2383	GATGATATATTTTAAGAGGCTPAAAAAATATATTAATCTACATATATGATTCGTTATGCTGATTC	2442
OY		1143	CTTTCTATTTTGTGATTAACAGGTTGCAGACACAGACAGAAAGTTTAAAGGATATAGATA	1202
Db		2443	CTTCTCATTTTGTGATTAACAGGTTGCAGACACAGACAGAAAGTTTAAAGGATATAGATA	2502
OY		1203	GTAAGAAAACAGATATGAGGATGATATCTTTTAAATCTTGGTGCAGAGCTTGGAAT	1262
Db		2503	GTAAGAAAACAGATATGAGGATGATATCTTTTAAATCTTGGTGCAGAGCTTGGAAT	2562
OY		1263	GACAGAGGCGCAAAGGAAAGAAACATTAATTTGATTATCAAAATATCATGATPATAG	1322
Db		2563	GACAGAGGCGCAAAGGAAAGAAACATTAATTTGATTATCAAAATATCATGATPATAG	2622
OY		1323	CAAAATGAAATATATTTTGTCCCAATTAAGAAATATAAATTTTCTCAAAACCAGATGG	1382
Db		2623	CAAAATGAAATATATTTTGTCCCAATTAAGAAATATAAATTTTCTCAAAACCAGATGG	2682
OY		1383	AAGCTCTACACAGCTTGAGGCAATACCTCTACATCCACATATAGAACGTTTCACTGAT	1442
Db		2683	AAGCTCTACACACAGCTTGAGGCAATACCTCTACATCCACATATAGAACGTTTCACTGAT	2742
OY		1443	ACACTTGTGCACAGGTTGGAGAGATGCTCAAGGATCCGAGTCCGAGATGACATTCACCAT	1502
Db		2743	ACACTTGTGCACAGGTTGGAGAGATGCTCAAGGATCCGAGTCCGAGATGACATTCACCAT	2802
OY		1503	GCGGCTTTTTGGGTCCACAGAGTACTCAAGACAGAGAGTACTACGGCTATGCTCGTTC	1562
Db		2803	GCGGCTTTTTGGGTCCACAGAGTACTCAAGACAGAGAGTACTACGGCTATGCTCGTTC	2862
OY		1563	GTCGNAAGNGTGATTCGCTGTAACTGTTGAAGTCTGTGATGTTGTGAAGCAAAACA	1622
Db		2862	GTCGNAAGNGTGATTCGCTGTAACTGTTGAAGTCTGTGATGTTGTGAAGCAAAACA	2922
OY		1623	GAGTCAGTTNGCTGATATCAAGGACTCCGAGCCAATCT	1660
Db		2922	GAGTCAGTTNGCTGATATCAAGGACTCCGAGCCAATCT	2959

[illegible]



**repeat\_region**

/note="corrected splice donor for PFC0005c"  
42416..43015  
/note="R-FA3"

**gene**

complement(join(43045..43971,44202..44255))  
/gene="PFC0010c, MAL3P8..2"

**CDS**

complement(join(43045..43971,44202..44255))  
/note="PFC0010c, MAL3P8..2"

aa; Similarity to other P.falciptarum rifins. Member of a multi-gene family; predicted using hexExon"

/codon\_start=1  
/protein\_id="CAB39116\_1"

/db\_xref="GI:4494021"

/translation="MKIHYINILFELPELNTLIYNQNHVITTPRPATRTPLCELEYAPSNDSPDEKRMWMOQVDPTDORFHEDYDERMTTRROCKEODKEIOKLILDKLEYSOMAOULTEPNITTEDIPTICOTESTADDETCNGCTGGVAPGAVIGSGLYVTNIVTVOALONCIENAGVEKAGEIELKGPGELIKLIKVSIOSFTIPANVAETTFENIRKYVENTKCVCHEHTCYSTYISKEMALSQRSELTAKNNADAKIAERVIDKAAPVTSGLNTAISIVIALVIIVLMIIITYLEFYRRNRKKRKRLDIYLILEE"

**gene**

complement(46448..47828)  
/gene="PFC0015c"

**CDS**

complement(46448..47828)  
/note="PFC0015c"

/pseudo

/codon\_start=1  
complement(join(48956..49777,49881..49949))  
/gene="PFC0025c, MAL3P8..4"

**gene**

complement(join(48956..49777,49881..49949))  
/note="PFC0025c, MAL3P8..4"

**CDS**

complement(join(48956..49777,49881..49949))  
/note="PFC0025c, MAL3P8..4"

stevor (3D7-stevortf3-1), len: 296 aa; Similarity to other P.falciptarum stevors. Member of a multi-gene family; predicted using hexExon"

/codon\_start=1  
/protein\_id="CAB39117\_1"

/db\_xref="GI:4494022"

/translation="MKNMYLLMLFEFLINTLIPHYEYTLNDYNSFTIONTKRTTIKSRLAQTQHINHPIHNDPELKETIDKMNEEAJTKYOQTHDPYEOLKEVENNKGTXYSGCNAAEISTELEKTLETYEIEFGNESDMKSGMSNPVBKSSFPFCTDINGKIAXITGRGRKYLLKHKGRCRTGRIGYCISGVITYITGTTAKGAIAISTAGSNCNICAIYSIFHLNLHKMSASLTRIALTGVCACADYASAGARSASAIFPCGIALVALVLITLVLLTYIWLYRRKRSMHECKRHLCX"

**gene**

complement(join(52066..53076,53192..53260))  
/gene="PFC0030c, MAL3P8..5"

**CDS**

complement(join(52066..53076,53192..53260))  
/gene="PFC0030c, MAL3P8..5"

note="PFC0030c (MAL3P8..5), rifin (3D7-rifrt3-2), len: 359 aa; Similarity to other P.falciptarum rifins. Member of a multi-gene family; predicted using hexExon"

/codon\_start=1  
/protein\_id="CAB39118\_1"

/db\_xref="GI:4494023"

/translation="WKLSYSNILFELPALNILLTSYVAHNKKKPSTPHHTRTTSTRLESCESTESIDSDELINSVKEIFEROJSRFEBYEREMTOROKRYEKQDKIOKLIHKDKMEIMLAKEIKEGCRCGCCGSVAGSIGLGCAVINMRPAALDAIAKITFGTANIAICAAVCRGEATGCVKLVISGLKKMGISITGLGNKESEPATSYKVVASNOAVECHKEECGEYGLKEGFPPGPSRDRIHCOSWOOTSUYSTGHGISPKVDIKYTEVMUSKAGPRNNAAEAYPKATEETRTIKAAEEKTETASTOLDYSALTIGISILAILIIVIMILIYLLIRKKRNKKMKKAQYTQLINE"

**gene**

join(58103..58174,58338..59390)  
/gene="PFC0035w, MAL3P8..6"

**CDS**

join(58103..58174,58338..59390)  
/gene="PFC0035w, MAL3P8..6"

note="PFC0035w (MAL3P8..6), rifin (3D7-rifrt3-3), len: 375 aa; Similarity to other P.falciptarum rifins. Member of a multi-gene family; predicted using hexExon"

/codon\_start=1  
/protein\_id="CAB39119\_1"

/db\_xref="GI:4494024"

/translation="WRITIKMKVHVINIILLATSLNIANTHOKPSSTPHHIQITRLL

Query Match	Best Local Similarity	Score	DB 3	Length	108908;
Matches	307;	Conservative	0;	Mismatches	293;
				Indels	5;
				Gaps	2;
gene					
CDS					
CECLYVHNNDNDPEKMRVQGHDRFTORFQYDERLDERLQKRYCVDCTDKLEKIL					
KDKLEKLNKFAFLQTDHSDAIPTICICTSMADYVEKTCIKTQNDGIVAPSSGV					
LAGIAEGALYAKMRFNALQTAIEALKNAADIDLVGGIEAGKRVVIGDGLAELEKGI					
GSMPDYFTEGYCINVKSLASIIYEKRTQGLCAKSTLDYATKSOISISIGTWMQDQTY					
GAPDPTFIEVLINGVEGTGEADVAAEARESEATSNKIVQERFLLEAGFNSSISSIN					
ASIIAIVIIILIMVIYIFETIRRYRKKIKKKILOYILLEE"					
Join(61248, .61301,61548, .62459)					
/gene="PFC0040W, MAL3P8.7"					
Join(61248, .61301,61548, .62459)					
/note="PFC0040W (MAL3P8.7), rifin, (3D7-rifT3-4), len: 322					
aa; Similarity to other P.f.alcaliprum rifins. Member of a					
multi-gene family; predicted signal and transmembrane					
sequence; predicted using hexexon					
/codon_start=1					
Query Match	6.4%;	Score 106.2;	DB 3;	Length 108908;	
Best Local Similarity	50.7%;	Pred. No. 6.1e-07;			
Matches	307;	Conservative	0;	Mismatches	293;
				Indels	5;
				Gaps	2;
47	TATTAAAGCCGACCTTAAACAAATGAGTGAAGAAAGAAAGAAATAATATATATATTT	106			
Db	39088 TATTGAAAGAAAGAAAGAAATGAAAGAAAGAAAGAAAGAAAGAAATATATATATAT	106			
QY	107 TGAAGTATGTAAGAGAGAAAGAAAGAAATATATATATATATATATATATATATAT	166			
Db	39148 TTATTTAAAT	166			
QY	167 AGAAT	226			
Db	39208 TAT	226			
QY	227 TTAATTTTCATTTTCGAGGATTAATTTGTTGTAAGTGTGATGAAAGAAATTCAGAT	286			
Db	39268 AAAAAAATTTTAT	286			
QY	287 TGCAGATCAAAA-ATGTTGTAATTTGACATTTATGAAATVACGTAGTATATCAATAT	345			
Db	39328 AAACCAAAAGAAAGAAAGAAATATATATATATATATATATATATATATATATAT	345			
QY	346 GGGTTTGTCTATTTTATTTTTCGCAAGTTGAAATTCGAGTGAAGAGAAATAGTTGG	405			
Db	39388 ATGAAAGAAATTTTATATATATATATATATATATATATATATATATATATAT	405			
QY	406 GAGCAAAAAAACCCCTGCGCTTTTTCGAATGACCTTTGGAAGAAATTCCTGTTG	465			
Db	39448 AAAAAAATTTTAT	465			
QY	466 TAGCGAAACCTTTGAAATTTTTCATTTGAAATTTTTCGAAGAAATTAAGCAAA	525			
Db	39504 AATATATGAT	525			
QY	526 AATGCAAAAAAT	585			
Db	39564 AAAAAAAT	585			
QY	586 TGACATTCATGAAAGAAAGAAATTTTCGAAGAAATTCATGCTGAGAGGATTTAA	645			
Db	39624 AAAAAAAT	645			
QY	646 AAAAA 650				
Db	39684 ATAAA 39688				
RESULT 4					
AX344555/c	349980 bp	DNA	linear	PAT 01-FEB-2002	
LOCUS	AX344555	Sequence 6 from Patent WO0200932.			
DEFINITION	AX344555				
ACCESSION	AX344555.1	GI:18492441			
VERSION					
KEYWORDS					
SOURCE					
synthetic construct.					



COMMENT On Aug 22, 2002 this sequence version replaced gi:21911559.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TrEMBL, Wp: WormPEP, Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP11-309M23 is from the library RPCR-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

FEATURES  
Source

1. 131682  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP11-309M23"  
/clone\_1lb="RPCR-11.2"  
BASE COUNT 34805 a 31378 c 30235 g 35264 t  
ORIGIN

Query Match 6.1%; Score 101.6; DB 9; Length 131682;  
Best Local Similarity 46.6%; Pred. No. 2.7e-06;  
Matches 360; Conservative 0; Mismatches 409; Indels 3; Gaps 1;

16 AAAATGAAAAATAGATTAATTTTCGACGATTTAAAGCCGACTTAAACCAATGAGTG 75

118367 AAAAAATGAATGAAATATGAAATTAATTAATTAATTAATTAATTAATTAATTA 118426

76 AACAAGAAAAAGAAAAATATACATATTTGAGTTAGTAAAGAAAGAAAAATTA 135

118427 AATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118486

136 AAGAGACCTCGGCTTAACAGTGAAGAAACCAAGAAATTAATAAAGAGAGCTGAT 195

118487 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118546

196 TTTTAATGGAATCGTGAAGAAAAAATTTTAAATTTTCATTTTCGAGGATTAATTG 255

118547 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118603

256 TTGTAGTGTATGAAAAATCTAGATTAATAAATGCAGATCAAAATGTGTTGAAATTGACA 315

118604 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118663

316 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 375

118664 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118723

376 AAAATCTGAGTGAAGAAATAGTTTGCAGAGCAAAAAAACCCCTTGGCTTTTTCAA 435

Db 118724 AATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118783

QY 436 ATGACTTTGCAAAAAATTCATTGTGACGGTAGCAAACTTTGAAATTTTTCATTGCA 495

Db 118784 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118843

QY 496 AATTGAAAAATTAAGCAAGAACTCAAAATGAAAAATTTATTAATTAATTAATTAATTA 555

Db 118844 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118903

QY 556 GATCGATATGATTTTAAAGCAAGAACTGACATTAATTAATTAATTAATTAATTAATTA 615

Db 118904 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 118963

QY 616 AATCTATGCTGAGAGTGAATTAAGCAAAAAATTTGGCACTCAACCCGTTGTTT 675

Db 118964 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 119023

QY 676 TAGCAATGACGATGAAGAAATGATGAAGAAATTAATTAATTAATTAATTAATTAATTA 735

Db 119024 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 119083

QY 736 CAGCTCAATATGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 787

Db 119084 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 119135

RESULT 6

AC087566

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

\*\*\*\*\*  
Project Information  
Center Project Name: 0  
Center clone name: RPCR-23\_127H3  
\*\*\*\*\*  
Summary Statistics  
Consensus quality: 150164 bases at least Q40  
Consensus quality: 175115 bases at least Q30  
Consensus quality: 186307 bases at least Q20  
Estimated insert size: 216300; agarose-fp estimation  
Estimated insert size: 226312; sum-of-contigs estimation  
Quality coverage: 3.5 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.34 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 57 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1074	contig of 1074	bp	in length
1075	1174	gap of unknown	length	
1175	2270	contig of 1096	bp	in length
2271	2370	gap of unknown	length	
2371	3511	contig of 1141	bp	in length
3511	3611	gap of unknown	length	
3612	4783	contig of 1172	bp	in length
4784	4883	gap of unknown	length	
4884	6377	contig of 1494	bp	in length
6378	6478	gap of unknown	length	
6478	7711	contig of 1234	bp	in length
7712	7811	gap of unknown	length	
7812	9092	contig of 1281	bp	in length
9093	9192	gap of unknown	length	
9193	10330	contig of 1138	bp	in length
10331	10430	gap of unknown	length	
10431	11677	contig of 1247	bp	in length
11678	11777	gap of unknown	length	
11778	13186	contig of 1409	bp	in length
13187	13286	gap of unknown	length	
13287	14565	contig of 1279	bp	in length
14566	14665	gap of unknown	length	
14666	15787	contig of 1122	bp	in length
15788	15887	gap of unknown	length	
15888	17012	contig of 1125	bp	in length
17013	17112	gap of unknown	length	
17113	18353	contig of 1241	bp	in length
18354	18453	gap of unknown	length	
18454	19611	contig of 1158	bp	in length
19612	19711	gap of unknown	length	
19712	21084	contig of 1373	bp	in length
21085	21184	gap of unknown	length	
21185	22398	contig of 1214	bp	in length
22399	22498	gap of unknown	length	
22499	23629	contig of 1131	bp	in length
23630	23729	gap of unknown	length	
23730	25112	contig of 1383	bp	in length
25113	25212	gap of unknown	length	
25213	26529	contig of 1317	bp	in length
26530	26629	gap of unknown	length	
26630	27773	contig of 1144	bp	in length
27774	27873	gap of unknown	length	
27874	28211	contig of 1338	bp	in length
28212	29311	gap of unknown	length	
29312	30438	contig of 1127	bp	in length
30439	30538	gap of unknown	length	
30539	31930	contig of 1392	bp	in length
31931	32030	gap of unknown	length	
32031	33112	contig of 1082	bp	in length
33113	33212	gap of unknown	length	
33213	34380	contig of 1168	bp	in length
34381	34480	gap of unknown	length	
34481	35515	contig of 1035	bp	in length
35516	35615	gap of unknown	length	
35616	37145	contig of 1530	bp	in length
37146	37245	gap of unknown	length	
37246	38655	contig of 1410	bp	in length
38656	38755	gap of unknown	length	
38756	40418	contig of 1663	bp	in length
40419	40518	gap of unknown	length	
40519	42089	contig of 1571	bp	in length
42090	42189	gap of unknown	length	
42190	43485	contig of 1296	bp	in length
43486	43585	gap of unknown	length	
43587	44557	contig of 1972	bp	in length
44558	44657	gap of unknown	length	
44658	47561	contig of 1904	bp	in length
47562	47661	gap of unknown	length	
47662	50270	contig of 2609	bp	in length
50271	50370	gap of unknown	length	
50371	52463	contig of 2093	bp	in length
52464	52563	gap of unknown	length	
52564	54241	contig of 1678	bp	in length

*	54242	54341: gap of unknown length
*	54342	56282: contig of 1941 bp in length
*	56283	56382: gap of unknown length
*	57724	57724: contig of 1342 bp in length
*	57725	57824: gap of unknown length
*	57825	60404: contig of 2580 bp in length
*	60405	60504: gap of unknown length
*	60505	63040: contig of 2536 bp in length
*	63041	63140: gap of unknown length
*	63141	64314: contig of 1174 bp in length
*	64315	64414: gap of unknown length
*	64415	68063: contig of 3649 bp in length
*	68064	68163: gap of unknown length
*	68164	72715: contig of 4552 bp in length
*	72716	72815: gap of unknown length
*	72816	79858: contig of 7043 bp in length
*	79859	79958: gap of unknown length
*	79959	87305: contig of 7347 bp in length
*	87306	87405: gap of unknown length
*	87406	94304: contig of 6899 bp in length
*	94305	94404: gap of unknown length
*	94405	98973: contig of 5569 bp in length
*	98974	100073: gap of unknown length
*	100074	110741: contig of 10668 bp in length
*	110742	110841: gap of unknown length
*	110842	116556: contig of 5715 bp in length
*	116557	116556: gap of unknown length
*	116557	126137: contig of 9481 bp in length
*	126138	126237: gap of unknown length
*	126238	139661: contig of 13424 bp in length
*	139662	139761: gap of unknown length
*	147923	147923: contig of 8162 bp in length
*	147924	148023: gap of unknown length
*	148024	160030: contig of 12007 bp in length
*	160031	160130: gap of unknown length
*	160131	175441: contig of 15311 bp in length
*	175442	175541: gap of unknown length
*	175542	191133: contig of 15592 bp in length
*	191134	191233: gap of unknown length
*	191234	231912: contig of 40679 bp in length

BASE COUNT	ORIGIN	clone_lib="RPC1 mouse BAC library 23"
61588	a	52985 c 59076 g 49983 t 8280 others

Query Match	6.18;	Score 101.6;	DB 2;	Length 231912;
-------------	-------	--------------	-------	----------------

Matches 433; Conservative 0; Mismatches 497; Indels 15; Gaps 2;

[illegible]

```

QY 321 GAATACGATATATCAATTAATGCGGTTTGTCTATTTATTTTCGAAGATTGAAT 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8318 AGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAA- 8376
QY 381 CTGAGTGAAGAAATATGTTTGGAGACAAAAAACCTTGCCGTTTTCCTCAATGAC 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8377 -AGAGGAGAAAGAAAAAATACAGCAAGAAAAAGAACAAAAAGAAAAAGAA 8435
QY 441 TTGGAAAAATTCATTGTGAGCGGTAGGAAACTTTGAATTTTTCATCTGGAATTT 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8436 AAAAAAGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8495
QY 501 GAAAAATAGGCAAGAACTCAATGAGAAAAATATTATTAATAAAAAAGAGATCG 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8496 AGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8555
QY 561 GATATGATTTTAAAGCAGAAACTGACATTCATGAAAAAAGATTTGGAAAAATC 620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8556 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8615
QY 621 TATGCTGAGGTGATTAAGCAAAAAATTTGGAACCTCAACCGGTGTTTATAGAA 680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8616 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8675
QY 681 ATGACGATGAAGAAATGATGAATAATATCACTGATGTTAATGAGAAACAGCAGT 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8676 AAAAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8735
QY 741 CATATAGCAATTTATTCAAAAATTAAGTTGACATAGTAATCAGATGATCTAGTACG 800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8736 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8795
QY 801 GGACT-----ATTAGAGTGGAACTGAACTGAACTGTTGATTAACAGAGAG 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8796 GAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8855
QY 848 TGCCTTTGCTTTTGTATTTGTGAGAGAAATTCAGCACTGAGAAAGAGCTGATATGC 907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8856 AGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8915
QY 908 AAGAGAGATTAAGATTTGATTCATGAAAGCAAAACTAAGA 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8916 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8960

RESULT 7
AC009524 174384 bp DNA linear HTG 07-JUN-2000
LOCUS Homo sapiens chromosome 15 clone RP11-31701 map 15q24, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC009524
AC009524.2 GI:8312269
VERSION HTG: HTGS_PHASE0.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 174384)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradaran,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A.,
Madan,A., Nesbitt,R., Shafer,T., and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 174384)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradaran,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Laskey,S.,
Madan,A., Ratcliffe,A., Shafer,T. and Hood,L.
COMMENT Submitted (26-AUG-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
On Jun 7, 2000 this sequence version replaced g1:5776572.
----- Genome Center
Center: Multimegabase Sequencing Center

```

```

Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_wmw
Contact: leierowensystembiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
-----
NOTE: This record contains 192 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
795 794: contig of 794 bp in length
895 894: gap of unknown length
1703 1702: contig of 808 bp in length
1803 1802: gap of unknown length
2637 2637: contig of 835 bp in length
2737 2737: gap of unknown length
3557 3557: contig of 820 bp in length
3657 3657: gap of unknown length
4453 4453: contig of 796 bp in length
4553 4553: gap of unknown length
5365 5365: contig of 812 bp in length
5466 5466: gap of unknown length
6259 6259: contig of 794 bp in length
6359 6359: gap of unknown length
7168 7168: contig of 809 bp in length
7268 7268: gap of unknown length
8063 8063: contig of 795 bp in length
8163 8163: gap of unknown length
8877 8877: contig of 814 bp in length
8977 8977: gap of unknown length
9078 9078: gap of 795 bp in length
9872 9872: contig of 795 bp in length
9873 9873: gap of unknown length
9972 9972: gap of 812 bp in length
10784 10784: contig of 812 bp in length
10785 10785: gap of unknown length
10884 10884: gap of 795 bp in length
11679 11679: contig of 813 bp in length
11779 11779: gap of unknown length
12592 12592: contig of 813 bp in length
12692 12692: gap of unknown length
13490 13490: contig of 798 bp in length
13590 13590: gap of unknown length
14404 14404: contig of 814 bp in length
14504 14504: gap of unknown length
15295 15295: contig of 791 bp in length
15395 15395: gap of unknown length
16208 16208: contig of 813 bp in length
16308 16308: gap of unknown length
17112 17112: contig of 804 bp in length
17212 17212: gap of unknown length
18028 18028: contig of 816 bp in length
18128 18128: gap of unknown length
18926 18926: contig of 798 bp in length
19026 19026: gap of unknown length
19839 19839: contig of 813 bp in length
19939 19939: gap of unknown length
20738 20738: contig of 799 bp in length
20838 20838: gap of unknown length
21653 21653: contig of 815 bp in length
21753 21753: gap of unknown length
22554 22554: contig of 801 bp in length
22654 22654: gap of unknown length
23468 23468: contig of 814 bp in length
23568 23568: gap of unknown length
24368 24368: contig of 800 bp in length

```

24369 24468: gap of unknown length  
24469 25282: contig of 814 bp in length  
25283 25382: gap of unknown length  
25383 26185: contig of 803 bp in length  
26186 26285: gap of unknown length  
26286 27100: contig of 815 bp in length  
27101 27200: gap of unknown length  
27201 28005: contig of 805 bp in length  
28006 28105: gap of unknown length  
28106 28921: contig of 816 bp in length  
28922 29021: gap of unknown length  
29022 29827: contig of 806 bp in length  
29827 29927: gap of unknown length  
29928 30736: contig of 809 bp in length  
30737 30836: gap of unknown length  
30837 31629: contig of 793 bp in length  
31630 31729: gap of unknown length  
31730 32554: contig of 825 bp in length  
32555 32654: gap of unknown length  
32655 33452: contig of 798 bp in length  
33453 33552: gap of unknown length  
33553 34362: contig of 810 bp in length  
34363 34462: gap of unknown length  
34463 35261: contig of 799 bp in length  
35262 35361: gap of unknown length  
35362 36271: contig of 810 bp in length  
36272 37067: contig of 796 bp in length  
37068 37167: gap of unknown length  
37168 37980: contig of 813 bp in length  
37981 38080: gap of unknown length  
38081 38674: contig of 794 bp in length  
38675 38974: gap of unknown length  
38975 39788: contig of 814 bp in length  
39789 39888: gap of unknown length  
39889 40686: contig of 798 bp in length  
40687 40786: gap of unknown length  
40787 41597: contig of 811 bp in length  
41598 41697: gap of unknown length  
41698 42496: contig of 799 bp in length  
42497 42596: gap of unknown length  
42597 43407: contig of 811 bp in length  
43408 43507: gap of unknown length  
43508 44317: contig of 810 bp in length  
44318 44417: gap of unknown length  
44419 45233: contig of 816 bp in length  
45234 45333: gap of unknown length  
45334 46127: contig of 794 bp in length  
46128 46227: gap of unknown length  
46228 47041: contig of 814 bp in length  
47042 47141: gap of unknown length  
47143 47937: contig of 796 bp in length  
47938 48037: gap of unknown length  
48038 48849: contig of 812 bp in length  
48849 48949: gap of unknown length  
48950 49747: contig of 798 bp in length  
49748 49847: gap of unknown length  
49848 50660: contig of 813 bp in length  
50661 50760: gap of unknown length  
50761 51559: contig of 799 bp in length  
51560 51659: gap of unknown length  
51660 52473: contig of 814 bp in length  
52474 52573: gap of unknown length  
52574 53375: contig of 802 bp in length  
53376 53475: gap of unknown length  
53476 54284: contig of 809 bp in length  
54285 54384: gap of unknown length  
54385 55187: contig of 803 bp in length  
55188 55287: gap of unknown length  
55288 56096: contig of 809 bp in length  
56097 56196: gap of unknown length  
56197 56992: contig of 796 bp in length  
56993 57092: gap of unknown length

57093 57907: contig of 815 bp in length  
57908 58007: gap of unknown length  
58008 58805: contig of 798 bp in length  
58806 58905: gap of unknown length  
58906 59718: contig of 813 bp in length  
59719 59818: gap of unknown length  
59819 59819: gap of unknown length  
60615 60614: contig of 796 bp in length  
60715 61524: contig of 810 bp in length  
61525 61624: gap of unknown length  
61625 62422: contig of 798 bp in length  
62423 62522: gap of unknown length  
62523 63332: contig of 810 bp in length  
63333 63432: gap of unknown length  
63433 64228: contig of 796 bp in length  
64229 64328: gap of unknown length  
64329 65138: contig of 810 bp in length  
65139 65238: gap of unknown length  
65239 66037: contig of 799 bp in length  
66038 66137: gap of unknown length  
66138 66948: contig of 811 bp in length  
66949 67048: gap of unknown length  
67049 67844: contig of 796 bp in length  
67845 67944: gap of unknown length  
67945 68757: contig of 813 bp in length  
68758 68857: gap of unknown length  
68858 69654: contig of 797 bp in length  
69655 69754: gap of unknown length

Query Match 6.0% Score 99.6: DB 2: Length 174384;  
Best Local Similarity 39.4%: Pred: No. 5e-06;  
Matches 535: Conservative 0: Mismatches 811: Indels 13: Gaps 3;

7 ACACAGCGCAATTTGCAAAATTTTCGACGATTAAGCCGCTTAAC 66  
103894 AGAAAAAATCTCTAAAAAATTAAGCAGCTGAAAAAGAAAAA 103953  
67 AAATGAGTGAAGAGAGAAAAAATTAATACATATTTTGAAGTAAAG 126  
103954 AAAGGAAGAGAAAAAACAATAATGAGAAAAAAGAAAGCTGAAT 104013  
127 AAAAAATTAAGAGACCTCGCTTAACAGTCGAAAAACCAATATAAAG 186  
104014 AAAAAAATAATACCGGAAAAAAGAAAAACAGAAAAATATACACA 104073  
187 GACTGTGATTTTAATGAAATCGAGAGAAAGAAATTTAATTTGATTTG 246  
104074 AAACGGAACAAAAAAGAGAGAGAAAAAAGAAAAACATACAAAAAG 104133  
247 ATTAATTTGTTGTAAGTGAAGAAATCTAGATTAATAATGCAATCA 306  
104134 AATAAT-----AAAAAGATTAANNNNNNNNNNNNNNNNNNNN 104187  
307 AATTGACATTAATGAAATAGCTAGTATATCAATAATGGGGTTTCTAT 366  
104188 NNN 104247  
367 CGAAGATTGAATACTGAGTGAAGAAATAGTTGCAGAGCAAAACCTGC 426  
104248 NNNNCGTTTTCACCGNATAGNTNAGAGTCGCGGNNNNNNNNNGCG 104307  
427 TTTTTCAAATGACTTTGGAATAATTCATTGTGAGCGGTAGCGAACT 486  
104308 AAGGAACCTTCAACCGAGCGCTGTATCCACCANNNNNNNNNTTAA 104367  
487 TACATTTGAATTTGAAATAATTAAGGCAAAAGCAATCGAAAAAATAT 546  
104368 AAAAAAGATTCCTTTAAAAAANNNNNNNNNNNNNNNNNNNNNNNN 104427  
547 AAAAAAGAGATCGGATTTGATTTTAAACAGCAAAATGACATTTGAAT 606  
104428 AAAAAAAGGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAN 104487



```

OY 607 ATTGGAAAAAATCTATGCTGAGCTGAATTAAGCAAAAAATTTGGAGCTCAACCG 666
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104488 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 104547
OY 667 GTGTTGTTTGAAGATGACGATGAAGAATATGAAAAATATCAACTCGATGTTATG 726
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104548 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAAAAAG 104607
OY 727 AAGAAACGACGCTCAATATAGCAATATTTCAAAAAATTAAGTGAAGCTAGTAATCG 786
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104608 AAAAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104667
OY 787 ATGATCTAGTAACGGGACTATTAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 846
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104668 GGGGGGNGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104723
OY 847 GTGCTCTTCTGTTTGTATTTGTGAGACATTCAGCACTGAGAGAGAGAGAGAGT 906
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104724 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104783
OY 907 CAAGAAGAGTAAAGATTTCATACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 966
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104784 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104843
OY 967 GTTTTGTGTTTGAAGATGATTTTGTAGTGAAGATGAATTCATGATGATTTCTC 1026
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104844 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104903
OY 1027 ACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104904 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104963
OY 1087 AATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104964 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 105023
OY 1147 CTATTGTTGTTGATTCAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 105024 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 105080
OY 1207 GAAACAGATATGAGATGATATCTTTTAAATATCTTGCTGCAAGAGTTGGA 1266
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 105081 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 105140
OY 1267 GAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 105141 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 105200
OY 1327 TGAATTAATATTTGTCGCAATTAAGAGATTAATATTTT 1365
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 105201 TTTTGAATATTTTATTAATTTTATTAATTTTATTAATTTT 105239

RESULT 8
AC079417 130540 bp DNA linear HTG 01-SEP-2000
LOCUS Mus musculus chromosome 16 clone RP23-102A24, WORKING DRAFT
DEFINITION
AC079417
AC079417.1 GI:9958029
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 130540)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
```

## COMMENT

```

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Project Name: 0
Center Project Name: 0
Center Clone name: RPCT-23_102A24
-----
Summary Statistics
Consensus quality: 71349 bases at least Q40
Consensus quality: 83788 bases at least Q30
Consensus quality: 89369 bases at least Q20
Estimated insert size: 93300: agarose-fp estimation
Estimated insert size: 127240: sum-of-ctdigs estimation
Quality coverage: 3.69 in Q20 bases; agarose-fp estimation
Quality coverage: 2.7 in Q20 bases; sum-of-ctdigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1043: contig of 1043 bp in length
1044 1143: gap of unknown length
1144 2160: contig of 1017 bp in length
2161 2260: gap of unknown length
2261 3696: contig of 1436 bp in length
3697 3797: gap of unknown length
3797 4826: contig of 1030 bp in length
4827 4926: gap of unknown length
4927 5959: contig of 1033 bp in length
5960 6059: gap of unknown length
6060 7579: contig of 1520 bp in length
7580 9178: gap of unknown length
9178 9278: contig of 1499 bp in length
9279 10501: gap of unknown length
10502 10601: contig of 1233 bp in length
10602 11610: gap of unknown length
11611 11710: contig of 1009 bp in length
11711 12840: gap of unknown length
12841 12941: contig of 1130 bp in length
12941 14262: gap of unknown length
14263 14362: contig of 1322 bp in length
14363 15984: gap of unknown length
15985 16084: contig of 1622 bp in length
16085 17502: gap of unknown length
17503 17602: gap of unknown length
17603 18902: contig of 1300 bp in length
18903 20008: gap of unknown length
20009 20108: contig of 1006 bp in length
20109 21372: gap of unknown length
21373 21472: contig of 1264 bp in length
21473 22675: gap of unknown length
22676 22775: contig of 1203 bp in length
22776 24088: gap of unknown length
24089 24188: contig of 1133 bp in length
24189 25354: gap of unknown length
25355 25454: gap of unknown length
25455 26641: contig of 1167 bp in length
26642 29340: gap of unknown length
29341 29441: contig of 2599 bp in length
29442 29440: gap of unknown length
29441 31609: contig of 2169 bp in length
31610 31709: gap of unknown length
31710 33695: contig of 1966 bp in length
33696 33795: gap of unknown length
33796 35117: contig of 1322 bp in length
```



CDS complement(1570..2424)  
 /gene="PFB0490C"  
 /note="predicted by Glimmerm"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAC71887.1"  
 /translation="MKENKIMDYLSCLPDVDVREKKSGKNSLTKSSSTKSDYKK  
 SIEFSKRDHSKSGSFRGRSGEINRKGSEPKPYNNRLIKNNYNNKGRPFHGR  
 DNEKGRSGSPGSPFNRRKSGEKRFRISNRKSSVSYGNGKNGKRSFNAPTR  
 TYVTKRLNNTKIVSAPVKKFNMLNISLRNRFALNTRSKFVGITIKSSVPKRIKK  
 GLKSGSLKSTRKSTGSKFKPLNKYFKLKIKIVTSINKIPSLKQKQKQEVNLPESL  
 NNATKKN"  
 11241..14606  
 /gene="PFB0495W"  
 11241..14606  
 /gene="PFB0495W"  
 /note="predicted by Glimmerm"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAC71888.1"  
 /DB\_xref="GI:3845198"  
 /translation="MFALKNTVREGFVNICFSYLKLYLKSNEVTNLYNETNNEKR  
 TNKKIKKSKAOSLFDKGLINHDKILFENLPYKCAKECISAKEVYKXLDYEKKC  
 FNYISLSDIIOSVKEIDELDKPTDYNEFYEVKNIDKNLKNELINEYFNKQDITFHR  
 ELIGKICNKIMSYIHEMGNELIHELIFEPNRNKNDKNLILFENYENYEDMTELEN  
 HEIKRLFLFNKTYLNNNSNIPPRKNLOEMERPLYEPREKNEKNTITKMKKEIYKK  
 CFAKHEVNDHIDNEKILNTRLYDNSLIDIDINKMLCNLNNNINENETIISKLL  
 NFYCTLIKRGKYDNDMTIYKLEVIKATHHILCDKRNKTECSDIDYSTLNSLNK  
 FLNKLIDKNFLIYFELKILKILNIFVNSQISLISLKNYILNRYNIVNV  
 LEFDIKFSLYCNIFLAGRIKTEENAVYLIHNODTVSNENKIDIIIOKRIEY  
 IFKMYENKDYHEFKLSDLSIKLSLSEVYKINEVNSYDFLLNNTSILYNNLY  
 NNSVKKYKDYIYIINDLSFYKIKKNDKREKKNFELSSKMEILCNILTSN  
 RYKIHHEEDNDQDQIVCSLFLNNLFEDKLIHPIYINLMCHYKTYNFKCKL  
 INEDIIISLLTCSKFQYFIENNSNDYRKELIHKYINIDILKNYLNYSISID  
 NISKFIISLNSKYCEVENBLLESIOSEFEKVTYSKGGIHMDDNNLIDNNNSCE  
 KYEHRIEYKKNELFENLKIIECLIKILYKKTLYIKOSLCPINLKNELI  
 KRLIYANNLYMYEMGYCENMERVLSKBEONLSYNNKNEKMDKILCHISE  
 DQYIEMSNMYIIFYDIKINSEROSLIRNSNDREPIDREKKEKYLNNNTLKH  
 NNKVLNEYKSNNSNGNISNLIKDDKRNHNEMDILDNKNEKKIOEGCONSEKEN  
 CDOVLNLIINIFGLKMERKKFLFOLYMYCNITKFRFRVSSSLPHMDVFKLIK  
 DMNLKLCIENRKINECAFVLTIDIVFEKER"

BASE COUNT 6284 a 1019 c 1106 g 6458 t

Query Match 5.8%; Score 96.6; DB 3; Length 14867;  
 Best Local Similarity 45.0%; Pred. No. 2.5e-05;  
 Matches 611; Conservative 0; Mismatches 729; Indels 17; Gaps 6;

QY 16 AATTTGAAAAATAGATAAATTTTCGAGGTATTAAGCCGATTAAACAATGATG 75  
 DB 9463 ATATTAAAAAGATATATTTTATTAATAATATGTTAAATATATAATATAG 9404  
 QY 76 AAGAGAGAAAAAGAAAAATTAATACATA-TTTTGAAGTATTAAGAGAAAAAATA 134  
 DB 9403 ATTAATTAATAAATTAATAAATATATTTTAAATATATATATATATATTA 9344  
 QY 135 AAGAGAGACCTGGCTTAACAGTCGAAAAACGAAATATTAAGAAGAAAG--AGACTG 191  
 DB 9343 AAAATATATATATTAATAAATTAATTAATTAATTAATAAATTAATAATAT 9284  
 QY 192 TGAATTTTATGAAATCGTAGAGAAAAATTTTAATTTTCATTTTCGAGGATTTAA 251  
 DB 9283 TAAATATTTAAATTAATTTAATATTAATTAATAAATTAATAATATATATA 9224  
 QY 252 TTTGTTGTAAGTGAAGAAAAATCTAGATAAAAATGACAGTCAAAAATGTTGAATT 311  
 DB 9223 AATAATATATATTAATAAATTAATAAATTAATAAATTAATAAATATATATAAT 9164  
 QY 312 GACATATTTGAATACGTATATATCAATTAATGGGGTTTGTCTATTTATTTTGGCAAG 371  
 DB 9163 AATAAATTAATAAATTAATAATTAATAAATTAATAATTAATTAATAAATATATAAT 9104

QY 372 ATTGAAATCTGAGTGAAAGAAAAATAGTTTGCAGAGCAAAAAACCGTTGCCGTTTTTT 431  
 DB 9103 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9044  
 QY 432 TCAATGACCTTGGAAAAAAATTCATGTGTGCGCGTACCGCAACTTGAATTTTTCAT 491  
 DB 9043 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8984  
 QY 492 TGAATTTGAAAAAATTAAGCAAGAAACCTCAATGAAAAAATTTTGTAAAAA 551  
 DB 8983 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8924  
 QY 552 AGAGATCGGATATGATTTTAAACAGAAAACTGCATTTGATGAAAAAAGATTG 611  
 DB 8923 TTTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8864  
 QY 612 GAAAAATCTATGCTGAGACTGAATTAAGCAAAAAAATTTGGAACTCAACCGGTGT 671  
 DB 8863 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8804  
 QY 672 GTTTTGAATATGACATGAAAGAAATGATGAAAAATATCAACCTCGATGTAAGA 731  
 DB 8803 ATTAATAATATAA--ATAATTAATATAAATAAATAAATAAATAAATAAATAAATAA 8746  
 QY 732 ACAGCAGCTCATATATGAAATTAATTCAAAAATAAAGTTGAGCATGTAATGATGAT 791  
 DB 8745 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8687  
 QY 792 CTAGTAACGGACTATTAAGAGTGTGAACCTGGAATAGTTTGTATTAACCAAGAGTCC 851  
 DB 8686 ATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8627  
 QY 852 TTTTCGTTTTGTATTTGTGTGAGAAATTCAGCAACTGAGAAAAAGAGTGAATGCAAG 911  
 DB 8626 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8567  
 QY 912 AGAGTAAGACATTTTCGATCAATGAAGCAAAAGCAATAAGAGGCTTTGATGAGTTT 971  
 DB 8566 ATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8507  
 QY 972 GTTTTGAATAGAGATTTTGTGAGTGAATTAAGATTAATCAATGATGATATTTTCAC 1031  
 DB 8506 ATATATTAATAA-----TAATGAAATTAATTAATTAATTTTATTAATTAATTAAT 8455  
 QY 1032 AAAAAAGCTCTGCAAGTAAAGAAAAACAATGAAGAGCGGACATGATGATTT 1091  
 DB 8454 TAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8395  
 QY 1092 TTTAAGAGCTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1151  
 DB 8394 AATTAATTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8335  
 QY 1152 TGTTCATTAACAGTTGACAGCAGCAGCAAGATTTTAAAGGATATAGATATGAAGAAC 1211  
 DB 8334 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8275  
 QY 1212 AGATATGAGATGATGATTTTAAATCTTGCGCAAGAGTTGGAATGACAGAGG 1271  
 DB 8274 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8217  
 QY 1272 CAAGAGCAAGCAATTAATTTGATTTATCAAAATATCATGATGATGAGCAATGAT 1331  
 DB 8216 AATTAATTAAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8157  
 QY 1332 TATATTTTGTGCAATTAAGATTAATTAATTTTCT 1368  
 DB 8156 TTTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8120

RESULT 10  
 AC096885 84563 bp DNA linear VRT 17-JAN-2002  
 LOCUS AC096885  
 DEFINITION Danio rerio clone 13614, complete sequence.

```

ACCESSION      AC096885
VERSION        AC096885.2  GI:18182756
KEYWORDS
SOURCE
ORGANISM
Danio rerio
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 84563)
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Jin,S.-O.,
Legaspi,R., Maduro,O.L., Maduro,V.B., Mastello,C., Maskeri,B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantrop,S., Thomas,J.W., Thomas,P.D., Touchman,D.W.,
Turgerson,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 84563)
Green,E.D.
Direct Submission
Submitted (29-SEP-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 84563)
Green,E.D.
Direct Submission
Submitted (17-JAN-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jan 17, 2002, this sequence version replaced gi:15808575.
----- Genome Center -----
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information -----
Center project name: agp
Center clone name: 136L04

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the Features section.

FEATURES
Source
1..84563
/organism="Danio rerio"
/db_xref="taxon:7935"
/clone="136L4"
/clone_1lb="Incyte Genomics"
48561..48672
/note="single clone coverage"

BASE COUNT      26959 a 15018 c 14630 g 27956 t

ORIGIN
Query Match      5.8%  Score 96.6;  DB 5;  Length 84563;
Best Local Similarity 45.5%;  Pred. No. 1.6e-05;
Matches 467;  Conservative 0;  Mismatches 549;  Indels 11;  Gaps 3;

32 TAAATTTTGCAGATTAATAAGCCGACTTAACAACAATGATGAGAGAGAAAGAAA 91
II IIII I I I I I I I I I I I I I I I I I I I I I I I I I
DB 47634 TATTAATTCCTGCTTTCTTAAATTAATAAATCAATTAATTAAGTAAGAAATTAACATAAAG 47699
ATTAATTAACATTAATTTGAGTTAGTAAAGAGAAAGAAAATTAATAAGAAAGACCTCGGCTT 151

```

Db	47694	ATATAATTAATAATTTTATTTATTAATAATAATAATTAATAATAATAATAATAATAA	47753
QY	152	AACAGTCGAAAAACGAGAAATAATTAAGAAAGAGACTGTCGATTTTAAATGGAATCGT	211
Db	47754	ATAAAATATTAATAAACCAATTAACATACTAAAAATAAAATATATTTTAAATTAATAAANA	47813
QY	212	GAGAAAGAAAGAAATTTTAAATTTTCATTTTCGAGGATTAATTTGTTGTAAGTTGATGTA	271
Db	47814	AATTAATAATAATAATAATAATGCTAAATAATGAATAAAAAAT-----TAAATAATAAANA	47867
QY	272	AATCTAGATTAATAAATGCAATGATCAAAATGTTGTAATTTGACATTTGTAATAATGCTAG	331
Db	47868	TAACTAATAATAATAATAATCAAAAAATAATCAATAAAAAATAATATACTAATTAATAANA	47927
QY	332	TAAATCAATAATGGGGGTTTGCTAATTTTATTTTTCGCAAGANTGAAAATCTGAGTGAAG	391
Db	47928	ACATAAATAATAATAATAATAATAATAATTAATAATGAATAATAATACAAATAATAATAACT	47987
QY	392	AAATATCTTTCGCGAGCAAAAAACCCTTCGCCGTTTTTTTCAATAGACTTTGGAAAAA	451
Db	47988	AGTCTAAATAATAATAATAATAATAATAATAATAATACTAAAAATAATAATAATAATAA	48047
QY	452	TTCAATTCGACGCGTACGCAACTTTGCAATTTTTCATATTTGCAATTTGAAAAATAAG	511
Db	48048	ATATTTAAATAATAATAATAATAATAATAATAATAATAATTCATTTAA---TAAAAAATAA	48104
QY	512	GCAAAAGAACTCAATGCAAAAATAATTTATTTTAAAAAAGAGATCGGATTTGCGATTT	571
Db	48105	AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	48164
QY	572	TAAAGAGAGAAATCGCATTTGATGAAAAAAGATTTGGAAAAAATCTATGCTGAGAG	631
Db	48165	TTTAATTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	48224
QY	632	TGATTTTAAAGCAAAAAAATTTGGGAACCTCAACCGGCTGTTTGAATGACGATGAA	691
Db	48225	TAAACACAGCTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	48284
QY	692	AGAAATGATGAAAAATATCACTCGATGTTTAATGAAGAAACGACGTCATATATAGAA	751
Db	48285	AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	48344
QY	752	ATTATTCAAAAAATTAAGTGAAGATAGTAATAGATGATCTGTGTACGCGACTATTACA	811
Db	48345	TTTTTTTTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	48404
QY	812	GTGTGAGACTCGAATATGTTTGAATAAACAAGACGCTTCGTTTGTGATTTGTGA	871
Db	48405	ATTAACATA--AATTAATATATAGGAAAAATAACAAAAACAATAATAATAATAATAA	48466
QY	872	GAGATTCAGCAACTGAGAAAGAGACGTGATTAATGCAAGAGAGTAAAAAGATTCGATAC	931
Db	48463	AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAC	48522
QY	932	AATGAAAGCAAAACCTAAGACGCTTTTCAATTTGACTTTTGTTTTGTATAGATTTTTT	991
Db	48523	TACTCTGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	48582
QY	992	GAGTGAATAATAGAAATCAATGGAATGATTTATTTTCAACAACAAAAAGACTGTCAGATA	1051
Db	48583	GAAATAAAAATAATAATGAATAAAACACTATATAATTAATAATAATAATAATAATAACTAT	48642
QY	1052	AAGAAAA 1058	
Db	48643	AATTAATA 48649	
RESULT	11		
E23356	E23356	6644 bp	DNA 1linear PAT 18-JUN-2001
DEFINITION	Virus vector system expressing apoptosis-related gene.		
ACCESSION	E23356		

VERSION	KEYWORDS	GI:13024379
JP 1999075859-A/2.	unidentified.	
SOURCE	unclassified.	
ORGANISM	unclassified.	
REFERENCE	1 (bases 1 to 6644)	
AUTHORS	Hirofumi, H.	
TITLE	Virus vector system expressing apoptosis-related gene	
JOURNAL	Patent: JP 1999075859-A 2 3-MAR-1999;	
COMMENT		
PC	PC	HIROFUMI HAMADA
PC	PC	C12N15/09, C12N5/10, C12N7/00, A61K35/76, A61K48/00, C12N5/10, PC
PC	PC	C12R1/91,
PC	PC	C12N7/00, C12R1/92, C12N15/00, C12N5/00, C12N5/00, C12R1/91
Strandedness:	Double;	
CC	Topology: Linear;	
FN	Key	Location/Qualifiers
FT	source	1. 6644
FEATURES		Location/Qualifiers
SOURCE		1. 6644
BASE COUNT	2166 a 1573 c 1424 g 1481 t	
ORIGIN		
Query Match	5.6% Score 93.6; DB 6; Length 6644;	
Best Local Similarity	45.9% Pred. No. 8.56-05;	
Matches 321; Conservative	0; Mismatches 379; Indels 0; Gaps 0;	
QY	7 ACACAGGCAATTCGAAAAATACATTAATTTTCGAGATTTAAAGCCGACTTAAC	66
DB	3787 AA	3846
QY	67 AATGAGTGAAGAGAAAGAAAGAAATTAATACATATTTTTCAGTGTAAAGAGAAAG	126
DB	3847 AA	3906
QY	127 AAAAAATPAAAGAAAGACCTCGGCTTAACAGTCGAAAAACGAAATATATAAAGAAAGA	186
DB	3907 AA	3966
QY	187 GACGTGATTTTAAATGGAATCGTGAAGAAAGAAATTTTAAATTTTCATTTTCGAGCG	246
DB	3967 AA	4026
QY	247 ATTAAATTTGTTGATGATGAAATATCTGATPAAATAATGCATCAAAATGTGTG	306
DB	4027 AA	4086
QY	307 AATTGACATTTTGAATACGTAATATATCAATTAATGAGGGTTGCTATTTTATTTTG	366
DB	4087 AA	4146
QY	367 CGAGATTTGAATATCTGAGTGAAGAAATATGTTTGGAGAGCAAAAAACCTTGCGCT	426
DB	4147 AA	4206
QY	427 TTTTTCGAATGACTTTGAAAAAATTCATTTGTGAGCGGTAGCGAACTTTGAATTTT	486
DB	4207 AA	4266
QY	487 TACATTTGGAATTTGAAAAAATAAGCAAAAGAACTCAATGGAATAATATTATTATA	546
DB	4267 AA	4326
QY	547 AAAAAAGAGATCGCATATGATTTTAAAGCAGAAAACTGACATTTGAATGAAAAAAG	606

Dd	4327	AAA	4386
Oy	607	ATTGTGAAAATAATCTAGTCGTGAAGTGAATTAAGAACAATAATGGAACTCAACCCG	666
Dd	4387	AAGATTTGGTGCAGACTTGCGC	4446
Oy	667	GTCGTGTTTTGGAATGACGATGGAAGAAATGATGAAAA	706
Dd	4447	GAGATTTTCAGAGACTAAGAAAGCTAAATTAATGGAGAAAAA	4486
 RESULT 12			
E23357	LOCUS	E23357	7372 bp DNA linear PAT 18-JUN-2001
DEFINITION	Virus vector system expressing apoptosis-related gene.		
ACCESSION	E23357		
VERSION	E23357.1 GI:13024380		
KEYWORDS	JP 1999075859-A/3.		
SOURCE	JP 1999075859-A/3.		
ORGANISM	Homo sapiens.		
	Homo sapiens.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
	1 (bases 1 to 7372)		
REFERENCE	Hirofumi,H.		
AUTHORS	Virus vector system expressing apoptosis-related gene		
TITLE	Patent: JP 1999075859-A 3 23-MAR-1999;		
JOURNAL	R B R JENSERU KK		
COMMENT	OS Homo sapiens (human) PN JP 1999075859-A/3 PD 23-MAR-1999 PF 08-SEP-1997 JP 1997259235 PR PC HIROFUMI HAMADA PC C12N15/09,C12N5/10,C12N7/00//A61K35/76,A61K48/00,(C12N5/10, PC C12R1:91), PC (C12N7/00,C12R1:92),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers FT source 1..7372 /organism='Homo sapiens (human)'. FT Location/Qualifiers 1..7372 /organism='Homo sapiens'		
FEATURES	Source		
BASE COUNT	2353 a 1749 c 1649 g 1621 t		
ORIGIN			
Query Match	5.6%, Score 93.6; DB 6; Length 7372;		
Best Local Similarity	45.9%; Pred. No. 8.3e-05;		
Matches 321; Conservative	0; Mismatches 379; Indels 0; Gaps 0;		
OY	7	ACACAAGCAATTTGAAAAAATAGATAAATTTTCGGCAGATTAAGCCGACTTAAC	66
Dd	4515	AAA	4574
OY	67	AAATGAGTGAAGAAGAAAGAAAAATTAATACATATTTTGACTTAGTAAGAAGAAAG	126
Dd	4575	AAA	4634
OY	127	AAAAAATBAAGAAGACCCTCGCTTAACAGTCGAAAAACGAAATTAATAAAGAAAGA	186
Dd	4635	AAA	4694
OY	187	GACGTGTATTTTAATGGAATCGTGGAGAAAGAAATTTTAATTTTCATTTTCGAGG	246
Dd	4695	AAA	4754
OY	247	ATTAAATTTGTTGATGATGAAGAAATCTAGATAAAAAATGCAGATCAAAATGTGTG	306
Dd	4755	AAA	4814
OY	307	AATTTGACATTATTTGAATACGTAGTATATCATATATGGGGTTTGCTATTTTATTTTG	366

[illegible][illegible]

RESULT 13					
LOCUS	E23355	7797 bp	DNA	linear	PAT 18-JUN-2001
DEFINITION	Virus vector system expressing apoptosis-related gene.				
ACCESSION	E23355				
VERSION	E23355.1	GI:13024378			
KEYWORDS	JP 1999075859-A/1.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 7797)				
AUTHORS	Hirofumi H.				
TITLE	Virus vector system expressing apoptosis-related gene				
JOURNAL	Patent: JP 1999075859-A 1 23-MAR-1999;				

D <sub>b</sub>	5420	AAA	5479
Q <sub>y</sub>	547	AAAAAAGACATCGATATGATTTTAAAGCAGAAAACGTGCATTGAATGAATAAAAAAG	606
D <sub>b</sub>	5480	AAA	5539
Q <sub>y</sub>	607	ATTTGAAATAATCTATGCTGAGAGTAAATTAATAAGCAAAAAATTGGAACTCAACCG	666
D <sub>b</sub>	5540	AAGATTGGTCGACCTTGGC	5599
Q <sub>y</sub>	667	GTCCTGTTTTGAAGAATGACCATGAAGAAGAAATGATGAAGAAA	706
D <sub>b</sub>	5600	GAGATTTTTCAGGAGCTAAGGAGCTAAATATCGAATAAAAA	5639

COMMENT	OS	Bovine poxvirus
	PN	JP 1999075859-A/1
	PD	23-MAR-1999
	PE	08-SEP-1997 JP 1997259235
	PR	
	PI	HIROFUMI HAMADA
	PC	C12N15/09,C12N5/10,C12N7/00,/A61K35/76,A61K48/00,(C12N5/10, PC C12N1:91), PC C12N7/00,C12R1:92),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC Strandedness: Double; CC Topology: Linear; CC Key key Location/Qualifiers 1..7797 FT source /organism='Bovine poxvirus'.
FEATURES		
source		location/Qualifiers 1..7797 /organism='unidentified' /db_xref='taxon:32644' 2542 a 1760 c 1656 g 1839 t
BASE COUNT		
ORIGIN		

RESULT 14	
LOCUS	E23359
DEFINITION	7996 bp DNA linear PAT 18-JUN-2001
VERSION	E23359
KEYWORDS	E23359.1 GI:13024382
SOURCE	JP 1999075859-A/5.
ORGANISM	Homo sapiens.
REFERENCE	Homo sapiens.
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1 (bases 1 to 7996)
	Hirofumi, H.
	virus vector system expressing apoptosis-related gene
	Patent: JP 1999075859-A 5 23-MAR-1999;
COMMENT	R B R JENSEN KK
	OS Homo sapiens (human)
	PN JP 1999075859-A/5
	PD 23-MAR-1999

	Query Match	5.6%	Score 93.6;	DB 6;	Length 7797;
	Best Local Similarity	45.9%;	Pred. No. 8,2e-05;		
	Matches	321;	Conservative	0;	Mismatches 379; Indels 0; Gaps 0;
QY	7 ACACAGCGCAATTCAGATAAATTTTCGCGAGTATTAAAGCCACTTAATC	66			
Db	4940 AA	49999			
YY	67 AAATGCTGAAGAAGAAAATAAATTAATACATATTTTGAGTTAGTAAGAAGAAG	126			

pr	HIROFUMI HAMADA	
PI	C12N15/09, C12N5/10, C12N7/00, A61K35/76, A61K48/00, C12N5/10, PC C12R1:911	
PC	(C12N7/00, C12R1:92), C12N15/00, C12N5/00, C12N5/00, C12R1:91) CC	
Stratadness:	Double;	
CC	Topology: Linear;	
FM	Key	Location/Qualifiers
FT	source	I. .7996

FEATURES FT /Organism='Homo sapiens (human)'  
 source 1..7996  
 /Organism='Homo sapiens'  
 /db\_xref='taxon:9606'  
 BASE COUNT 2463 a 2015 c 1829 g 1689 t  
 ORIGIN

Query Match 5.6%; Score 93.6; DB 6; Length 7996;  
 Best Local Similarity 45.9%; Pred. No. 8.1e-05;  
 Matches 321; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

QY 7 ACACAGGCGCAATTTGAAATAGATAATTTCCAGCGTATTAAGCGCATTAAC 66  
 Db 5139 AAA 5198  
 QY 67 AATGAGTGAAG 126  
 Db 5199 AAA 5258  
 QY 127 AAAAAAAAAAG 186  
 Db 5259 AAA 5318  
 QY 187 GACTGTGATTTTATGAAATCGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
 Db 5319 AAA 5378  
 QY 247 ATTAAATTTGTTAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
 Db 5379 AAA 5438  
 QY 307 AATTGACATTTATGAAATCTAGATATCATATATGAGGCTTTGCTATTTATT 366  
 Db 5439 AAA 5498  
 QY 367 CGAGATTTGAAATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
 Db 5499 AAA 5558  
 QY 427 TTTTTCGAAATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
 Db 5559 AAA 5618  
 QY 487 TACATGGAATTTGAAATTAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
 Db 5619 AAA 5678  
 QY 547 AAAAAAAAAAGAGATGAGATGAGATTTTAAAGAGAGAGAGAGAGAGAGAGAG 606  
 Db 5679 AAA 5738  
 QY 607 ATTGGAG 666  
 Db 5739 AAA 5798  
 QY 667 GTCTGTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 706  
 Db 5799 GAGATTTTCAG 5838

RESULT 15  
 AC116966/c 268147 bp DNA linear HTG 04-APR-2002  
 LOCUS AC116966  
 DEFINITION Dictyostellium discoideum chromosome 2 map 2482102-2750247 strain  
 AX4, \*\*\* SEQUENCING IN PROGRESS \*\*\*; in ordered pieces.  
 ACCESSION AC116966  
 VERSION AC116966.1 GI:19920065  
 KEYWORDS HTG: HTGS PHASE2.  
 SOURCE Dictyostellium discoideum.  
 ORGANISM Dictyostellium discoideum  
 Eukaryota; Eukaryota; Dictyostellida; Dictyostellium.  
 REFERENCE 1 (bases 1 to 268147)

AUTHORS Gloeckner, G., Eichinger, L., Szafarski, K., Pacheco, J., Dear, P.,  
 Lehmann, R., Baumgart, C., Parra, G., April, J.F., Guigo, R., Kumpf, K.,  
 Tungal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A. and  
 Noegel, A.A.  
 TITLE Sequence and Analysis of Chromosome 2 of Dictyostellium  
 JOURNAL Unpublished  
 REFERENCE The Dictyostellium Genome Sequencing Consortium  
 2 (bases 1 to 268147)  
 AUTHORS Baumgart, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-APR-2002) Genome Analysis, Institute of Molecular  
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
 COMMENT CDS predictions from GenemID may contain errors. Further information  
 is available from IMB Jena, Department of Genome Analysis  
 (http://genome.imb-jena.de/dictyostellium/)  
 and the University Cologne, Institute for Biochemistry I  
 (http://www.uni-koeln.de/dictyostellium/project.shtml)  
 Funding  
 Agency : Deutsche Forschungsgemeinschaft (DFG).  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source 1..268147  
 /Organism='Dictyostellium discoideum'  
 /strain='AX4'  
 /db\_xref='taxon:44689'  
 /Chromosome='2'  
 /map='2482102-2750247'  
 BASE COUNT 101633 a 30111 c 30461 g 105740 t 202 others  
 ORIGIN

Query Match 5.6%; Score 93.4; DB 2; Length 268147;  
 Best Local Similarity 45.5%; Pred. No. 3.6e-05;  
 Matches 409; Conservative 0; Mismatches 486; Indels 4; Gaps 2;

QY 6 TACACAGCGCAATTTGAAATAGATAATTTCCAGCGTATTAAGCGCATTAAC 65  
 Db 197289 TAA 197230  
 QY 66 CAATGAGTGAAG 125  
 Db 197229 AAA 197170  
 QY 126 GAAAAAAAAAG 185  
 Db 197169 AAA 197110  
 QY 186 AGACTGATTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245  
 Db 197109 AAAAAAAAAACCTTTTGTGAGCTGCTTTATTAATTAATTAATTAATTT 197050  
 QY 246 GATTAAATTTGTTAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305  
 Db 197049 TTTGAGCTTATTTATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 196990  
 QY 306 GAATTTGACATTTATGAATACGATATATCATATATGAGGCTTGTATTTAT 365  
 Db 196989 --ATTCATATGTTTAAATTCACACACCTCTAATTTGTTAAATTAAGAGAAA 196933  
 QY 366 GCGAGATTTGAATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
 Db 196932 ATAA 196873  
 QY 426 TTTTTCGAATGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485  
 Db 196872 AAAAAAAAAAG 196813  
 QY 486 TTACATTTGAG 545  
 Db 196812 TATCACTATTAAGCTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 196753

OY	546	AAAAAAAAAGGAGTTCGATGTGATTTAAACGAGAAAACTGCATTTGATGATGAAAAAAA	605
Db	196752	AAAAAAAAAAAAAAAAAGAAACAACAAAAAAAAAAAAAAAAAAAAAAAAATCATTAAGAAAAAAA	1966939
OY	606	GATTTCGAAAAAATCATGTCTGAGAGTGAAATTAAGCAAAAAAATTTGGAACTCAACC	665
Db	196692	AAAAAAAAAATATC-ATTAAAAAATATTTTCCAAAGTCAAACAAATTAAGTTCACCACT	196634
OY	666	GCTGTGTTTAAAGATGACGATGAAAGAAATGATGAAAAATTAACAACCTCGATGTTAAT	725
Db	196633	TAATGATTACTCTAATAGAAATTAATTAATTTTCAACACTTATACACACACAATTA	196574
OY	726	GAAGAACAGCAGCTCATATATGGAATATTTCAAAATTAAGTTGACATAGTAAATCA	785
Db	196573	TTACAAATTAATTTTCATTTCCACCAAGAAAAAATTAATTAATTAACCAAGATTAT	196514
OY	786	GATGATTAGTAACGGACACTTACAGTCTGGAACTCGAATATAGTTTGTATTAACAAGA	845
Db	196513	AAATTAATAAAAAAAAAAATTAATAAAATGTCAACACACCAAAATATATATATAATTAAG	196454
OY	846	AGTCGCTTCGTTTTGTATTTGTAGAGAAATTCAGCACTCGAAGAAAGAGCTGATTA	904
Db	196453	TAAGTATTTTCATCTTTAAAAAATAAAAAAAAAAAAAAAAAAATTAATTAAGTAAATTAATTA	196395

Search completed: July 17, 2003, 04:54:12  
Job time : 4319.71 secs